

General Economic Principles of Bargaining and Trade: Evidence from 2,000 Classroom Experiments

Po-Hsuan Lin
National Taiwan University and MobLab
pohsuanlin.0105@gmail.com

Alexander L. Brown
Texas A&M University
alexbrown@tamu.edu

Taisuke Imai
LMU Munich
taisuke.imai@econ.lmu.de

Joseph Tao-yi Wang
National Taiwan University
josephw@ntu.edu.tw

Stephanie W. Wang
University of Pittsburgh
swwang@pitt.edu

Colin F. Camerer*
California Institute of Technology
camerer@hss.caltech.edu

September 15, 2018

Abstract

Standardized classroom economics experiments provide a treasure trove of evidence about how well results reproduce when nearly-identical methods are used. We use a sample of around 20,000 observations to test reproducibility in bargaining and trading. Ultimatum bargaining exhibits some geographical variation, and shows that equal split offers are accepted more often and more quickly than slightly unequal offers. Double auction results are highly reproducible and are close to equilibrium predictions. Our large sample shows robust, strong correlations between how individual surplus and trading order, and autocorrelation (-0.5) of successive price changes, consistent with trading dynamics based on limited rationality.

Keywords: Generalizability, ultimatum game, double auction, price formation, heterogeneity
JEL codes: C88, D44, C78

One-sentence summary: In hundreds of classroom experiments, buyer-seller market results are the same, but two-person bargaining outcomes vary.

*To whom correspondence should be addressed. We thank Mathew O. Jackson, John O. Ledyard, John Duffy and Ming Hsu for helpful feedback and the MobLab engineering team for technical assistance. Imai was supported by Deutsche Forschungsgemeinschaft through CRC TRR 190. J. T.-y. Wang was supported by CRETA, National Taiwan University (NTU-107L900203; MOST 107-3017-F-002-004) and the Ministry of Science and Technology of Taiwan (MOST 106-2628-H-002-001-MY4).

Economists and other social scientists generally search for lawful regularities in human behavior, and the causes of those regularities. A prerequisite for lawful regularity is that when identical methods are used by different scientists the results are roughly the same. However, testing for regularity in this way is extremely challenging because even the most careful attempted replications inevitably differ in some ways from original studies ([Open Science Collaboration, 2015](#); [Camerer et al., 2016, 2018](#)).

To learn about the extent of regularity and difference in simple economic systems, we use data from an educational technology platform (called MobLab). The platform offers highly standardized interfaces and data collection. Software standardization is prized by instructors because it makes teaching simpler. Standardized experimentation should also be prized by scientists because it fixes the details of methods, to generate nearly-identical replications, except for differences associated with different student demographics and classroom experiences ([Zwaan et al., 2018](#)).

We explore experimental regularity in two important domains of collective human behavior—two-person bargaining, and centralized buyer-seller trade. The large data set enables us to draw strong conclusions about how bargaining and centralized trade differ. It has been hypothesized that two-person bargaining is “personal,” activating evolved concerns for fairness and reciprocity from our ancestral life ([Smith, 1998](#)). These concerns could generate differences in bargaining outcomes across different geographical settings, and even across types of classes and students due to cultural norms about sharing and classroom exposure.

Centralized trade, in contrast, is thought to be “impersonal”: A market of people acting selfishly, with no concern for fairness, can produce mutually-beneficial outcomes. As Adam Smith famously said, “not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard to their own interest” ([Smith, 1776](#)). Indeed, economic theory has shown mathematically why concerns for fairness which influence two-person bargaining will have no impact on large markets ([Sobel, 2009](#); [Dufwenberg et al., 2011](#)). This sweeping hypothesis suggests centralized buyer-seller trading experiments may produce nearly identical outcomes for students all over the world.

Our data provide the most comprehensive test of regularity of behavior in experimental bargaining and centralized trading using an identical protocol in many countries. The two major limits of this approach are that subjects are all students in classes (in a narrow range of age and analytical skill), and most of them are not incentivized by financial rewards, as is the norm in contemporary experimental economics. In early studies ([Chamberlin, 1948](#); [Smith, 1962](#)), however, subjects were not incentivized for trades, but those unincentivized data were important in suggesting basic principles about behavior in markets. It later became customary to pay subjects earnings linked to experimental behavior, and to even pay a small “commission” for executing trades (since otherwise marginal trades were not made; [Plott and Smith, 1978](#)). Summaries of different experiments with within-experiment variation in no, medium, and high performance incentives have sometimes shown decreased variability in response noise with incentive pay ([Smith and Walker, 1993](#); [Her-](#)

twig and Ortmann, 2001), although the most reliable differences from not paying come when a response is socially desirable (e.g., taking more risk, or sharing money with strangers; Camerer and Hogarth, 1999). Attempts to conduct the ultimatum game without monetary incentives, such as Tompkinson and Bethwaite (1995), also find roughly similar results as when incentives are used Güth et al. (1982). While we believe that incentive pay linked to choices is usually desirable in economics experiments, the experimental platform simply did not impose this on instructors.

1 The Data

MobLab is an online educational platform for conducting economics experiments. Instructors sign up on the MobLab website¹ and invite students to participate in experiments from a large menu. Instructors use the platform so that their students can participate personally in simple versions of the kinds of games and markets the students are reading about in textbooks. The ultimatum game and buyer-seller market are two of the most popular experiments. The popularity of these two games gives ample observations (nearly 10,000 plays of each) and wide geo-demographic variation.

In each session of the ultimatum game, student players are randomly assigned to be a Proposer or a Responder. The Proposer offers a share of a fixed amount (a “pie size”) to the Responder. The Responder can accept the offer or reject it; if it is rejected both players get zero. The actual screens that subjects see are shown in Figures 1A and 1B. For each session, the instructor can specify the number of periods each pair plays together, and the pie size. The most common pie size is the default setting, which is 100 units (used in 82% of pairs). The modal number of plays by each pair is 1 and the average is 1.615, and we have a total of 10,507 observations (Figure 1E).

In every market session of the double auction, student players are divided randomly into two subgroups, buyers and sellers. Each buyer (seller) gets information about the private values (costs) of the goods that they can trade before entering the market. Figures 1C and 1D show their screens. Values and costs are induced artificially (Smith, 1976). For example, a buyer might be told that if they buy a unit they can redeem it for, say, \$3. A seller might be told they can keep the money they sell a unit for minus a cost of \$1. In those examples, the buyer and seller should trade at a price between \$1 and \$3; if they do trade they create a “surplus” of \$2. The players can use a slider bar to post a bid to buy one unit of the good (or post an ask to sell), which is then displayed for all players to see. Players may also trade by clicking a button to buy or sell at a current bid or ask price. To limit mistakes, if a player chooses a potential bid or ask that brings negative profit (based on their induced value or cost), the slider bar turns red as a warning. Instructors can specify any market configuration of values and costs they like. However, the default configuration (described in Section 3) is the most common setting (used in 36% of the markets). We have a total of 9,023 observations from 5,809 markets for the double auction (Figure 1F), after excluding one market session due to extremely low efficiency.²

¹<https://www.moblab.com/>

²See Appendix B for details of this market session. Including this market session obviously affects the efficiency

2 Ultimatum Games

The ultimatum game is a simple form of take-it-or-leave-it bargaining (Güth et al., 1982; Camerer, 2003; Oosterbeek et al., 2004). If all players care only about their own payoffs, Responders should accept any positive proposal offer (and maybe even zero). Proposers who anticipate that behavior by Responders should offer the smallest amount (or zero). As has been observed in many previous studies, this extreme prediction is far from what the student players (and other human subjects) commonly do.

Some Responders reject positive offers, even though they know they will get less money. There does not appear to be strong learning over repeated plays of ultimatum games (even against the same partner) to accept offers in general. There is evidence of a tendency to accept higher offers and reject lower offers more frequently over time (Cooper and Dutcher, 2011).

The most prominent explanation for rejection of offers is that Responders are willing to sacrifice earnings to reduce disadvantageous inequality, which results if they get less than the Proposer (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). However, a distaste for inequality does not seem to be sole explanation because Responders are more likely to reject small offers when they are made by a human Proposer, rather than by a computer or disinterested third-party (Blount, 1995), indicating negative reciprocity toward people who have treated them unfairly (Rabin, 1993; Falk and Fischbacher, 2006).

Selfish Proposers who anticipate Responders' rejections of small offers should make strategic offers that are large enough to avoid rejection. Proposers might also make large generous offers if they are altruistic, although most evidence suggests Proposers are approximately maximizing, given expectations of rejection by inequality-averse or reciprocal Responders (Forsythe et al., 1994; Levine, 1998).

Figure 2A and 2B shows distributions of Proposer offers, and acceptance rates for different offers in our sample. The distribution of offers shows clear multimodal spikes at multiples of tens, most frequently at 50%. This result echoes with past results showing that focal points play a significant role of determining the bargaining outcomes (Roth, 1985). The acceptance rate of Responders increases almost monotonically with the size of the proposal offer (except for a slight downward trend for those rare hyperfair offers of more than 50%).

One feature of behavior in these ultimatum games is that there is something special about offers of exactly 50% (this is only statistically clear because of the large volume of data). A piecewise linear regression of acceptance probability as a linear function of offers, with a break at 50%, shows a highly significant discontinuous jump in acceptance at 50% in both the one-shot game and repeated games (one-shot 16.3%, $t = 10.84$, $p\text{-value} < 0.001$; repeated 26.2%, $t = 9.40$, $p\text{-value} < 0.001$). The jump is robust to fitting a quadratic acceptance rate curve (see Appendix A.1).³

results, but not the other results.

³The discontinuous drop immediately *after* 50% is insignificant in both one-shot games and repeated games with $p\text{-value} = 0.188$ and 0.557 , respectively. This is in contrast to Backus et al. (2018) who find equal-split offers treated

Another feature our data can show is the effect of conducting experiments in different classes. Figure 2C shows the regression coefficients for class dummies when predicting proposal offers controlling for various configurations (see Appendix A.3). We find that only advanced elective classes in economics significantly reduce proposal offers (students in these classes are also likely to major in economics). In game theory and political economy classes, proposer offers are lower by more than 5% (from an average of 40.04%). In contrast, in experimental and behavioral economics and management/finance classes proposal offers are lower by only 2–3.2%. Our results are consistent with Bauman and Rose (2011) who find lower charity donation caused by self-selection for students majoring in economics, not but “indoctrination” of non-majors in those classes. While our data cannot show whether more economics training reduces proposal offers because of self-selection or causality, it, nevertheless, shows a correlation.

The MobLab platform also records how long it takes subjects to make decisions (their reaction times, or RTs). The Responders’ RTs indicate the special appeal of equal-split offers: there is a drastic drop in the time Responders take to make a decision when they face an offer of 50%, say, compared to 49% (see Figure 2D and Appendix A.4). For one-shot games, Responders’ average RT drops from 13.5 seconds to 8.9 seconds when they are facing an equal-split offer compared to 49% offers. As acceptance rates increases proportionally to offer size, we also find an inverted-U shape in RT as acceptance rate increases from 40% to 100%, peaking at 55-60%. This relationship shows that, at the population level, the Responders’ RT locally peaks at where they are indifferent with accepting and rejecting the offer and drops at where they have the highest acceptance rate. This pattern is consistent with a growing body of literature about inferring preferences from RT (Chabris et al., 2009; Konovalov and Krajbich, 2017). Furthermore, our population-level result is consistent with Krajbich et al. (2014), though their discrete-offer design is not capable of detecting the sharp drop at exactly 50%.

The frequency of exactly 50% splits, and the speed and jump in acceptance, is not consistent with typical models of inequity aversion or reciprocity—giving 49% rather than 50%, for example, should not lead to a jump in rejections. However, models with social image concerns can produce equilibria with statistical concentrations around an exact 50-50 norm. This type of model is consistent with our data, and is consistent with other data on dictator lab experiments (Andreoni and Bernheim, 2009) and the frequency of exactly equal sharing in bequests (Bernheim and Severinov, 2003) and in invention royalty sharing (Kotha et al., 2018).

Roth et al. (1991) (hereafter RPOZ) is an early study which explored differences in ultimatum games across four countries. Appendix A.2 compares offer distributions and acceptance in their four countries and in our sample. The general patterns are similar, but they do not detect a discontinuous jump in acceptance rate for 50% offers (the jump is only 4.5%).

differently in alternating-offer bargaining.

3 Buyer-Seller Double Auctions

Centralized markets, in which buyers and sellers compete impersonally, are the major building blocks of theoretical microeconomics. Such markets also represent how much of the trading in developed economies is actually organized.

Some of the earliest economic experiments showed that even with a small numbers of traders on each side of the market, traders can arrive at mutually beneficial trades as long as there is centralized exchange of information about prices and trading opportunities—called “price discovery” (Smith, 1962; Plott and Smith, 1978). The key idea used to analyze these experiments is the familiar “competitive equilibrium” (CE). The CE price is the price at which the number of units bought equals the number sold (the intersection of supply and demand curves). In the markets created by MobLab, in CE the traders are collectively maximizing the sum of profit (“producer’s surplus”) and consumer surplus, which is the total net value of goods purchased minus prices paid.

Figure 3A shows the most common supply-demand design in our dataset. It has five buyers and five sellers who can each trade up to three units.⁴ The CE price range is [110, 114] with a midpoint of 112, and generates 13 units of trade. Note that these CE price and quantity predictions are based on *no* free behavioral parameters: They rely only on the behavioral assumptions that no traders think they can manipulate the price by themselves (called “price taking”), and on some imagined process of price adjustment leading to CE. The CE predictions could easily be wrong for many reasons: Traders might prefer to trade at round numbers, they might trade too impulsively or wait too long, or buyer and seller groups could, in principle, collude among themselves to withhold trade and manipulate prices. Furthermore, as we note in more detail below, standard proofs about CE generally rely on the assumption that there are large numbers of traders on both sides of the market. There is no proof that convergence to CE will occur with small numbers of traders, who may strategically anticipate price-setting by others.

Figure 3A shows 10 time series of trade prices for four consecutive trading periods, one from each geographical region. The average price is fairly close to the predicted CE, but the dynamic process varies across different markets. Although the dynamics are all different, two general features arise: First, the volatility of price change decreases in the later periods. Second, regardless of initial transactions, the last price seems indicative of the equilibrium price.

In Figure 3B, we analyze the relation between price volatility and allocation efficiency.⁵ We

⁴In our buyer-seller double auction, each bid or ask is only for a single unit of the commodity, which is different from multiple unit double auction that the players can trade multiple units at one time.

⁵Efficiency is defined as $E \equiv (\sum_{i=1}^N \pi_i) / (\sum_{i=1}^N \pi_i^C)$, where N is the number of commodities can be traded, π_i is the realized profit and π_i^C is the CE profit of good i .

measure price volatility with Smith's alpha,⁶ and find an average of 0.278 (S.D. = 0.294), ranging from 0.136 in China to 0.393 in Canada, while community college students yield 0.534. Efficiency is 81.5% (S.D. = 0.258) ranging from 72.7% in England to 86.0% in Spain, although community college students achieve only 69.5%. Conceptually, if the price fluctuates dramatically, it is possible that some traders who should not trade that unit in equilibrium can transact at a profit under extreme price levels. Either the buyer/seller on the other end of the trade is trading at a loss, or this transactor will crowd out another who should trade in equilibrium, decreasing efficiency. Subsequently, there should be a negative correlation between Smith's Alpha and efficiency (see Figure 3B), though this can only be shown in large data sets with many markets. Indeed, the correlation coefficient is -0.079 (p-value < 0.001) for our data. When considering repeated markets after more than three periods of trading ($n = 469$ from 317 markets), the correlation coefficient is much stronger at -0.409 (p-value < 0.001). This indicates that experience with the same supply-demand schedule helps players identify and exploit arbitrage opportunities during price fluctuations. Fitting an exponential convergence model proposed by Gjerstad (2007) (see Appendix B.2) predicts Smith's alpha dropping from 0.206 to 0.086 if students were to play 25 periods, but efficiency remains stable at about 92% across periods.

We compare our results to Ketcham et al. (1984) to examine the consistency of our results to previously published laboratory experimental studies. Their Design I consists of four buyers and four sellers who each have a demand or supply of three units and repeatedly trade in the same continuous double auction market for 25 periods. They report an efficiency level of 95.89% (in periods 7 to 9) and an asymptotic Smith's alpha of 5.9%. In our data, these numbers are estimated to be 92.08% (in period 7) and 8.6% (in period 25, their last period).

Although the data show that most of the available surplus can be harvested by the double auction, efficiency losses are still common across markets. To analyze the sources of inefficiency, we adopt the same approach as Cason and Friedman (1996) and decompose the efficiency losses to two different types: transactions involving extra-marginal trades (*EM-inefficiency*) and profitable trades that are not realized (*V-inefficiency*). In our data, 47.4% of the losses are from extra-marginal trades and the rest are from unrealized profitable trades. In Cason and Friedman (1996), 55.8-67.8% of the loss is from extra-marginal trades. This implies both *EM-inefficiency* and *V-inefficiency* are prominent, though the higher rate of unrealized trades in our classroom data, compared to Cason and Friedman (1996), may be due to the lack of financial stakes.

Within a period, in order to quantify the heterogeneity in the initial prices and the convergence of ending price, we adopt the convergence model developed by Noussair et al. (1995):

$$y_{it} = (1/t)\mathbb{X}_i \cdot \beta_1 + (1 - 1/t)\beta_2 + \epsilon_{it},$$

⁶The Smith's alpha measures the price change volatility in a market of Q transactions, and is defined as

$$\alpha \equiv \frac{\sqrt{\frac{1}{Q} \sum_{q=1}^Q (P_q - P_C)^2}}{P_C},$$

where P_C is the equilibrium price and P_q is the actual price of each transaction. This was first proposed by Smith (1962), who also set P_C to be the midpoint of the CE price range.

where i indicates the particular period of market, t represents time as measured by the transaction order, \mathbb{X}_i is a vector containing the information about that market period, and β_1 explains the origin of the possible dynamic process. Therefore, β_2 is the asymptote of the dependent variable. As t becomes larger, the weight of β_1 becomes smaller because $1/t$ approaches 0 while the weight of β_2 increases as $(1 - 1/t)$ approaches to 1. We take the standardized price difference as the dependent variable.⁷ Figure 3C is the coefficient plot and the full regression table is in Appendix B.1. Although the initial price is on average 15.3% higher than the equilibrium price with subtle heterogeneity, the final price converges to only 1.7% higher than the equilibrium price, within the range of -2.9% to 1.9% reported in Ketcham et al. (1984).

Figure 3D provides a scatter plot (in log-scale) for the actual observed quantity and the predicted equilibrium quantity within each of the market experiments. Each dot represents one experimental trading period. An ordinary least squares regression line has a slope that is close to the identity trend line ($R^2 = 0.760$, slope $\hat{\beta} = 0.967$ and clustered S.E. = 0.012), but with a significant downward shift of 10.0% (p-value = 0.004). There is a little less trading than predicted (which is common in early periods of previous incentivized lab experiments also), Smith (1962) also finds slightly less trading without monetary incentives and show how real incentives and small trade commissions induce subjects to bid up to their values or costs.

The CE predicted price and quantity are rather accurate predictions of actual prices and quantity. However, it has long been mysterious how the psychology and dynamics of trading over real time actually leads to these market outcomes. Three prominent theories have been proposed for within-period price formation, but have only been tested in small samples. In addition, Easley and Ledyard (1993) propose a theory for between-period price formation, but also could not test it very powerful due to data limitations. The volume of our data provides a powerful test of all these important theories.

The three within-period dynamic theories range widely over how strategically sophisticated players are assumed to be. From most sophisticated to least, one theory assumes Bayesian mutual adjustment of trading strategies in an incomplete information game (Wilson, 1987, hereafter “Mutual Adjustment”), a second theory simplifies by assuming traders do not strategize about other traders’ strategies (Friedman, 1991, hereafter “Against Nature”), and a third theory proposed by Gode and Sunder (1993) posits “zero-intelligence” traders who pay no attention to price history at all and just bid randomly (while not losing money). See Appendix B.3 for details.

The three theories predict that the auto-correlation between successive price *changes* within a trading period should be either zero, positive, or negative, respectively. The first two theories also predict a clear correlation between how extreme a trader’s values or costs are, and when they trade: The buyers and sellers who have the most to earn are predicted to trade the earliest. This prediction corresponds to a negative correlation of value and trade order for buyers, and a positive

⁷The standardized price difference is defined as $\delta_q \equiv (P_q - P_C)/P_C$, where P_C is the equilibrium price and P_q is the actual price of each transaction with $q \in \{1, \dots, Q\}$.

correlation of cost and trade order for sellers. Thus, all three theories can be tested by looking at auto-correlation of price changes and correlation of value/cost and trade order.

Figures 4A and 4B show the distributions of Spearman's rank correlation coefficients of the buyer value and order of trade, and seller cost and order of trade. Each market period with at least four transactions contributes one correlation.⁸ The average rank-order correlation coefficient is $\rho_{\text{Buyer}} = -0.539$ (S.D. = 0.277) for buyers and $\rho_{\text{Seller}} = 0.518$ (S.D. = 0.279) for sellers. Hence, unlike Cason and Friedman (1996) which find weak evidence, we find stronger support for the two strategic theories (Wilson, 1987; Friedman, 1991), though zero-intelligence theory also suggests high-value buyers and low-cost sellers will transact slightly earlier.

In addition, given the large volume of data, we are able to identify a decrease in the size of this correlation when comparing period 1 to 3 with period 4 and beyond. In fact, buyer rank-order correlation drops from $\rho_{\text{Buyer}}^{1-3} = -0.540$ (S.D. = 0.277) to $\rho_{\text{Buyer}}^{\text{other}} = -0.506$ (S.D. = 0.275), passing the Kolmogorov-Smirnov test with p -value = 0.006 ($KS = 0.0826$). Similarly, seller rank-order correlation decreases from $\rho_{\text{Seller}}^{1-3} = 0.521$ (S.D. = 0.278) to $\rho_{\text{Seller}}^{\text{other}} = 0.472$ (S.D. = 0.299) with p -value = 0.003 ($KS = 0.0867$). This provides the first large-scale evidence for the theory of between-period price formation developed in Easley and Ledyard (1993).

Figure 4C is the scatter plot of price change from one trade to the next, plotted against the lagged price change. Similar to the inexperienced sessions of Cason and Friedman (1996), there is a clear trade-by-trade negative auto-correlation, estimated to be -0.457 ($R^2 = 0.209$, $\hat{\beta} = -0.491$ with clustered S.E. = 0.007, $\hat{\alpha} = -0.007$ with clustered S.E. = 0.001), which is very close to the prediction of -0.50 from the zero-intelligence theory and far from the prediction of the two other theories.

Figure 4D reports all three correlations at once, showing a three-dimensional scatter plot in the top-right along with the three-way view for buyer rank-order correlation, seller rank-order correlation, and price change auto-correlation. Since there is no clear prediction for zero-intelligence algorithm, we run 10,000 simulations in our default market to create a benchmark. Matching our data with the 95% confidence region of zero-intelligence simulations and two other theoretical predictions, we find much stronger support for zero-intelligence theory, compared to earlier evidence of Cason and Friedman (1996) and Cliff and Bruten (1998). However, a non-negligible portion of our data falls outside of the 95% confidence region, pointing toward the other two more sophisticated theories of price formation.

4 Heterogeneity

With data from around the globe, we are able to look into heterogeneity generated by geographical differences. Specifically, we are interested in whether there is country variation in proposal offers and acceptance rates in the ultimatum game, as well as in first and last transaction prices in

⁸This is because we need four transactions to calculate price change auto-correlation, leaving us with 8,492 markets.

buyer-seller double auctions. In order to control for the experience and environment, we use the data from the first period of the default configuration in the following analysis.

Conceptually, we want to analyze how subject pools differ in our data set. To quantify how an *average* player/market would vary across different regions, we focus on the variation of “true means” in different countries, instead of the variation in each player/market’s performance. To achieve this goal, we treat each country’s data set as a different “study” (that estimates its country mean) and conduct meta-analysis.

Specifically, we calculate the I^2 proposed by Higgins et al. (2003) through the DerSimonian-Laird random-effects model, commonly used in meta-analyses (DerSimonian and Laird, 1986):

$$I^2 = \left(\frac{Q - df}{Q} \right) \times 100\%$$

where df is the degree of freedom and Q is the weighted sum of squares (of each country’s deviation from the true mean) weighted by the inverse-variance.⁹ If there is no heterogeneity across different countries, df is the expected weighted sum of squares. Therefore, since Q is the observed weighted sum of squares, I^2 represents the fraction of variation attributed to between-country instead of within-country heterogeneity (see Appendix C).

We find cross-country variation only for proposal offers in the ultimatum game, with 86.6% of the heterogeneity coming from between-country heterogeneity. Figure 5A shows that, compared to the average offer of 35, students in Spain make significantly lower offers of 25–30, while students in Canada and South/West US make slightly more generous offers of 36–37.¹⁰ In contrast, acceptance rates in the ultimatum game vary insignificantly across countries. Figure 5B shows that only 30.6% of the heterogeneity stems from between-country variation, with no country yielding a significantly different acceptance rate outcome. RPOZ also find cross-country variation in proposal offers, and less variation in corresponding acceptance rates, although tests of both offers and acceptance do not have strong power. They speculate this is due to culture differences in social norms regarding acceptable offers that are well-understood and respected in each culture, so that offers vary more than acceptances do. However, they acknowledge that their “conjecture must stand or fall on the repeatability and robustness of these results” (p. 1092). Using thousands of classroom experiments, we are able to demonstrate the repeatability and robustness of the same results in the ultimatum game, as well as the universal effect of competition in the buyer-seller double auction. In fact, our

⁹Here we adopt the random-effect model, which assumes the true means for different countries are heterogeneous. Moreover, the Q statistic is defined as

$$Q = \sum_{k=1}^K w_k \left[m_k - \frac{\sum_{k=1}^K w_k m_k}{\sum_{k=1}^K w_k} \right]^2 \text{ for } w_k = \frac{1}{s_k^2}$$

where m_k and s_k are the sample mean and its standard error of country k , and K is the number of countries. See Higgins and Thompson (2002) and appendix C for the details of estimation.

¹⁰However, all ultimatum game observations from Spain come from game theory classes, which itself induces low proposal offers. Splitting the data, we find students in non-game theory classes have $I^2 = 56.97\%$, while those in game theory classes result in $I^2 = 74.01\%$. In other words, more than half of the heterogeneity comes between-country heterogeneity for both game theory and non-game theory classes.

results of the double auction are even more robust, and exhibit little between-country difference in both initial and final transaction prices (Figures 5C, 5D). In other words, we find little between-country differences even for initial prices that are potentially influenced by culture differences.

5 Discussion

We live in a golden age of social science. There is a symbiotic growth in available data and in development of methods for figuring out what the data tell us. The best new data sets have volume, variety, and velocity. In our case, volume and variety of experimental economics data show some interesting novel findings, and differences.

Our analysis demonstrates that experimental creations of two fundamental, but different, kinds of economic activity—outcomes in two-person ultimatum bargaining and 10-person buyer-seller trading—are closely reproducible in a large sample of classroom experiments. The behavior in buyer-seller trading, in particular, should be considered as reproducible as the kinds of experiments that are done in a college chemistry lab in order to demonstrate to students the regularity of chemical reactions. In fact, regularly run classroom experiments serve as another venue of replication that repeats these classic economic experiments over and over, distinguishing them from well-known experiments of other social sciences that have sometimes failed to replicate (Camerer et al., 2016, 2018; Ebersole et al., 2016; Klein et al., 2014, forthcoming).

At the same time, Variability across classroom groups is consistent with the hypothesis that bargaining is more personal, and hence has more variable outcomes, than impersonal competitive market trading. This is not too surprising as local norms of fair sharing are expected to affect outcomes. Indeed, games like these are now commonly used to study cross-cultural differences precisely because they are thought to reflect sharing norms, and perhaps influences such as marketization and religion (Henrich, 2000; Henrich et al., 2010a, 2005).

While there are between-country differences in bargaining offers and acceptances, differences in market trading are very small. Indeed, the results from simple competitive buyer-seller trading appear to be as close to a culturally universal, highly reproducible outcome as one is likely to get in social science about collective behavior. This bold claim is limited, of course, by the fact that all these data are high school and college students in classes in “WEIRD” societies (Henrich et al., 2010b). Given this apparent robustness, it would next be useful to establish if emergence of CE in small buyer-seller markets extends to small-scale societies, across the human life cycle, to adult psychopathology and cognitive deficit, and even to other species.

One new observation is that in ultimatum game bargaining, three large differences occur for offers of exactly half (compared to offers slightly below or above). At that offer level, there is a sharp jump in offer frequency, a discontinuous jump in acceptance rates, and an unusual drop in reaction times by Responders. The frequency of 50% offers has often been noted (Forsythe et al., 1994; Camerer and Thaler, 1995; Roth, 1997), but the other two effects on acceptance rate and

speed have not been established before.

Another important empirical observation we are able to make with substantial confidence is the regularity of dynamics within a trading period, in buyer-selling trading markets. There is a robust association between value and cost magnitudes and trading order, and a negative correlation between successive price changes close to -0.50 . Earlier exploratory studies have shown these patterns, but without anywhere near the degree of confidence that comes from the volume of data here. (For example, the price change auto-correlations are seen in every country and are extremely close in magnitude; they range from -0.459 to -0.415 .) These two empirical facts are accounted for by a simple theory of random zero-intelligence trading. However, that theory makes a lot of other counterfactual predictions—viz., there is no scope for learning across periods, so the simple ZI model cannot explain the between-period convergence (as [Easley and Ledyard \(1993\)](#) do). So there is room for improvement by adding more intelligence to that theory. Our data provide raw evidence that can be immediately used to test any such theory, with extremely high statistical power.

Besides the advantages which come from having a large volume and variety of data, the *velocity* of new data accumulation is high. Broad conclusions such as ours can therefore be re-evaluated rapidly, in years rather than decades. Educational platforms like MobLab which are continuously churning out new (and standardized) data, and welcome improvements on all sorts of openness and data sharing cheered on by journals, funding agencies and everyone else, will continue to make more and better data available in the years ahead. Our analysis is one glimpse of how such data can make the foundations of important aspects of social science sturdier and sturdier.

References

- Andreoni, James and B. Douglas Bernheim**, “Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects,” *Econometrica*, 2009, 77 (5), 1607–1636.
- Backus, Matthew, Thomas Blake, Bradley Larsen, and Steven Tadelis**, “Sequential Bargaining in the Field: Evidence From Millions of Online Bargaining Interactions,” 2018. NBER Working Paper No. 24306.
- Bauman, Yoram and Elaina Rose**, “Selection or Indoctrination: Why Do Economics Students Donate Less than the Rest?,” *Journal of Economic Behavior & Organization*, 2011, 79 (3), 318–327.
- Bernheim, B. Douglas and Sergei Severinov**, “Bequests as Signals: An Explanation for the Equal Division Puzzle,” *Journal of Political Economy*, 2003, 111 (4), 733–764.
- Blount, Sally**, “When Social Outcomes Aren’t Fair: The Effect of Causal Attributions on Preferences,” *Organizational Behavior and Human Decision Processes*, 1995, 63 (2), 131–144.
- Bolton, Gary E. and Axel Ockenfels**, “ERC: A Theory of Equity, Reciprocity, and Competition,” *The American Economic Review*, 2000, 90 (1), 166–193.
- Camerer, Colin F.**, *Behavioral Game Theory: Experiments in Strategic Interaction*, Princeton, NJ: Princeton University Press, 2003.

- **and Richard H. Thaler**, “Anomalies: Ultimatums, Dictators and Manners,” *Journal of Economic Perspectives*, 1995, 9 (2), 209–219.
- **and Robin M. Hogarth**, “The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework,” *Journal of Risk and Uncertainty*, 1999, 19 (1-3), 7–42.
- **, Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Johan Almenberg, Adam Altmejd, Taizan Chan et al.**, “Evaluating Replicability of Laboratory Experiments in Economics,” *Science*, 2016, 351 (6280), 1433–1436.
- **, – , Felix Holzmeister, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Gideon Nave, Brian A. Nosek, Thomas Pfeiffer et al.**, “Evaluating the Replicability of Social Science Experiments in *Nature* and *Science* between 2010 and 2015,” *Nature Human Behavior*, 2018, 2, 637–644.
- Cason, Timothy N. and Daniel Friedman**, “Price Formation in Double Auction Markets,” *Journal of Economic Dynamics and Control*, 1996, 20 (8), 1307–1337.
- Chabris, Christopher F, Carrie L Morris, Dmitry Taubinsky, David Laibson, and Jonathon P Schuldt**, “The allocation of time in decision-making,” *Journal of the European Economic Association*, 2009, 7 (2-3), 628–637.
- Chamberlin, Edward H.**, “An Experimental Imperfect Market,” *Journal of Political Economy*, 1948, 56 (2), 95–108.
- Cliff, Dave and Janet Bruten**, “Less Than Human: Simple Adaptive Trading Agents for CDA Markets,” in “Proceedings of the 1998 Symposium on Computation in Economics, Finance, and Engineering: Economic Systems” 1998.
- Cooper, David J. and E. Glenn Dutcher**, “The Dynamics of Responder Behavior in Ultimatum Games: A Meta-Study,” *Experimental Economics*, 2011, 14 (4), 519–546.
- DerSimonian, Rebecca and Nan Laird**, “Meta-Analysis in Clinical Trials,” *Controlled Clinical Trials*, 1986, 7 (3), 177–188.
- Dufwenberg, Martin, Paul Heidhues, Georg Kirchsteiger, Frank Riedel, and Joel Sobel**, “Other-Regarding Preferences in General Equilibrium,” *The Review of Economic Studies*, 2011, 78 (2), 613–639.
- Easley, David and John O. Ledyard**, “Theories of Price Formation and Exchange in Double Oral Auctions,” in Daniel Friedman and John Rust, eds., *The Double Auction Market: Institutions, Theories, and Evidence*, Addison-Wesley, 1993, pp. 63–97.
- Ebersole, Charles R., Olivia E. Atherton, Aimee L. Belanger, Hayley M. Skulborstad, Jill M. Allen, Jonathan B. Banks, Erica Baranski, Michael J. Bernstein, Diane B. V. Bonfiglio, Leanne Boucher et al.**, “Many Labs 3: Evaluating Participant Pool Quality across the Academic Semester via Replication,” *Journal of Experimental Social Psychology*, 2016, 67, 68–82.
- Falk, Armin and Urs Fischbacher**, “A Theory of Reciprocity,” *Games and Economic Behavior*, 2006, 54 (2), 293–315.
- Fehr, Ernst and Klaus M. Schmidt**, “A Theory of Fairness, Competition, and Cooperation,” *The Quarterly Journal of Economics*, 1999, 114 (3), 817–868.

- Forsythe, Robert, Joel L. Horowitz, Nathan E. Savin, and Martin Sefton**, “Fairness in Simple Bargaining Experiments,” *Games and Economic Behavior*, 1994, 6 (3), 347–369.
- Friedman, Daniel**, “A Simple Testable Model of Double Auction Markets,” *Journal of Economic Behavior & Organization*, 1991, 15 (1), 47–70.
- Gjerstad, Steven**, “The Competitive Market Paradox,” *Journal of Economic Dynamics and Control*, 2007, 31 (5), 1753–1780.
- Gode, Dhananjay K. and Shyam Sunder**, “Allocative Efficiency of Markets With Zero-Intelligence Traders: Market as a Partial Substitute for Individual Rationality,” *Journal of Political Economy*, 1993, 101 (1), 119–137.
- Güth, Werner, Rolf Schmittberger, and Bernd Schwarze**, “An Experimental Analysis of Ultimatum Bargaining,” *Journal of Economic Behavior & Organization*, 1982, 3 (4), 367–388.
- Henrich, Joseph**, “Does Culture Matter in Economic Behavior? Ultimatum Game Bargaining Among the Machiguenga of the Peruvian Amazon,” *The American Economic Review*, 2000, 90 (4), 973–979.
- , **Jean Ensminger, Richard McElreath, Abigail Barr, Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas, Michael Gurven, Edwins Gwako, Natalie Henrich et al.**, “Markets, Religion, Community Size, and the Evolution of Fairness and Punishment,” *Science*, 2010, 327 (5972), 1480–1484.
- , **Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, Richard McElreath, Michael Alvard, Abigail Barr, Jean Ensminger et al.**, ““Economic Man” in Cross-Cultural Perspective: Behavioral Experiments in 15 Small-Scale Societies,” *Behavioral and Brain Sciences*, 2005, 28 (6), 795–815.
- , **Steven J. Heine, and Ara Norenzayan**, “The Weirdest People in the World?,” *Behavioral and Brain Sciences*, 2010, 33 (2-3), 61–83.
- Hertwig, Ralph and Andreas Ortmann**, “Experimental Practices in Economics: A Methodological Challenge for Psychologists?,” *Behavioral and Brain Sciences*, 2001, 24 (3), 383–403.
- Higgins, Julian P. T. and Simon G. Thompson**, “Quantifying Heterogeneity in a Meta-Analysis,” *Statistics in Medicine*, 2002, 21 (11), 1539–1558.
- , —, **Jonathan J. Deeks, and Douglas G. Altman**, “Measuring Inconsistency in Meta-Analyses,” *BMJ: British Medical Journal*, 2003, 327 (7414), 557.
- Ketcham, Jon, Vernon L. Smith, and Arlington W. Williams**, “A Comparison of Posted-Offer and Double-Auction Pricing Institutions,” *The Review of Economic Studies*, 1984, 51 (4), 595–614.
- Klein, Richard A., Kate A. Ratliff, Michelangelo Vianello, Reginald B. Adams Jr., Štěpán Bahník, Michael J. Bernstein, Konrad Bocian, Mark J. Brandt, Beach Brooks, Claudia Chloe Brumbaugh et al.**, “Investigating Variation in Replicability: A “Many Labs” Replication Project,” *Social Psychology*, 2014, 45 (3), 142–152.
- , **Michelangelo Vianello, Fred Hasselman, Byron Gregory Adams, Reginald B. Adams Jr., Sinan Alper, Diego Vega, Mark Aveyard, Jordan Axt, Mayowa Babaloia et al.**, “Many Labs 2: Investigating Variation in Replicability across Sample and Setting,” *Advances in Methods and Practices in Psychological Science*, forthcoming.

- Kononov, Arkady and Ian Krajbich**, “Revealed indifference: Using response times to infer preferences,” 2017. SSRN No. 3024233.
- Kotha, Reddi, Young-Choon Kim, Colin Camerer, and Dan Lovallo**, “Equality Norms, Incentives, and Value Claiming in Scientific Licensing Contracts,” 2018. Unpublished manuscript.
- Krajbich, Ian, Bastiaan Oud, and Ernst Fehr**, “Benefits of Neuroeconomic Modeling: New Policy Interventions and Predictors of Preference,” *The American Economic Review*, 2014, *104* (5), 501–06.
- Levine, David K.**, “Modeling Altruism and Spitefulness in Experiments,” *Review of Economic Dynamics*, 1998, *1* (3), 593–622.
- Noussair, Charles N., Charles R. Plott, and Raymond G. Riezman**, “An Experimental Investigation of the Patterns of International Trade,” *The American Economic Review*, 1995, *85* (3), 462–491.
- Oosterbeek, Hessel, Randolph Sloof, and Gijs van de Kuilen**, “Cultural Differences in Ultimatum Game Experiments: Evidence From a Meta-Analysis,” *Experimental Economics*, 2004, *7* (2), 171–188.
- Open Science Collaboration**, “Estimating the Reproducibility of Psychological Science,” *Science*, 2015, *349* (6251), aac4716.
- Plott, Charles R. and Vernon L. Smith**, “An Experimental Examination of Two Exchange Institutions,” *The Review of Economic Studies*, 1978, *45* (1), 133–153.
- Rabin, Matthew**, “Incorporating Fairness into Game Theory and Economics,” *The American Economic Review*, 1993, *83* (5), 1281–1302.
- Roth, Alvin E.**, “Toward a Focal Point Theory of Bargaining,” in Alvin E Roth, ed., *Game-theoretic Models of Bargaining*, Cambridge, UK: Cambridge University Press, 1985, pp. 259–68.
- , “Bargaining Experiments,” in John Kagel and Alvin E Roth, eds., *Handbook of Experimental Economics*, Vol. 1, Princeton University Press, 1997, pp. 253–348.
- , **Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir**, “Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study,” *The American Economic Review*, 1991, *81* (5), 1068–1095.
- Smith, Adam**, *An Inquiry Into the Nature and Causes of the Wealth of Nations*, Vol. 1, London, UK: W. Strahan and T. Cadell, 1776.
- Smith, Vernon L.**, “An Experimental Study of Competitive Market Behavior,” *Journal of Political Economy*, 1962, *70* (2), 111–137.
- , “Experimental Economics: Induced Value Theory,” *The American Economic Review*, 1976, *66* (2), 274–279.
- , “The Two Faces of Adam Smith,” *Southern Economic Journal*, 1998, *65* (1), 2–19.
- **and James M. Walker**, “Monetary Rewards and Decision Cost in Experimental Economics,” *Economic Inquiry*, 1993, *31* (2), 245–261.
- Sobel, Joel**, “Generous Actors, Selfish Actions: Markets With Other-Regarding Preferences,” *International Review of Economics*, 2009, *56* (1), 3–16.

- Tompkinson, Paul and Judy Bethwaite**, “The Ultimatum Game: Raising the Stakes,” *Journal of Economic Behavior & Organization*, 1995, 27 (3), 439–451.
- Wilson, Robert B.**, “On Equilibria of Bid-Ask Markets,” in George R. Feiwel, ed., *Arrow and the Ascent of Modern Economic Theory*, Springer, 1987, pp. 375–414.
- Zwaan, Rolf A., Alexander Etz, Richard E. Lucas, and M. Brent Donnellan**, “Making Replication Mainstream,” *Behavioral and Brain Sciences*, 2018, 41, e120.

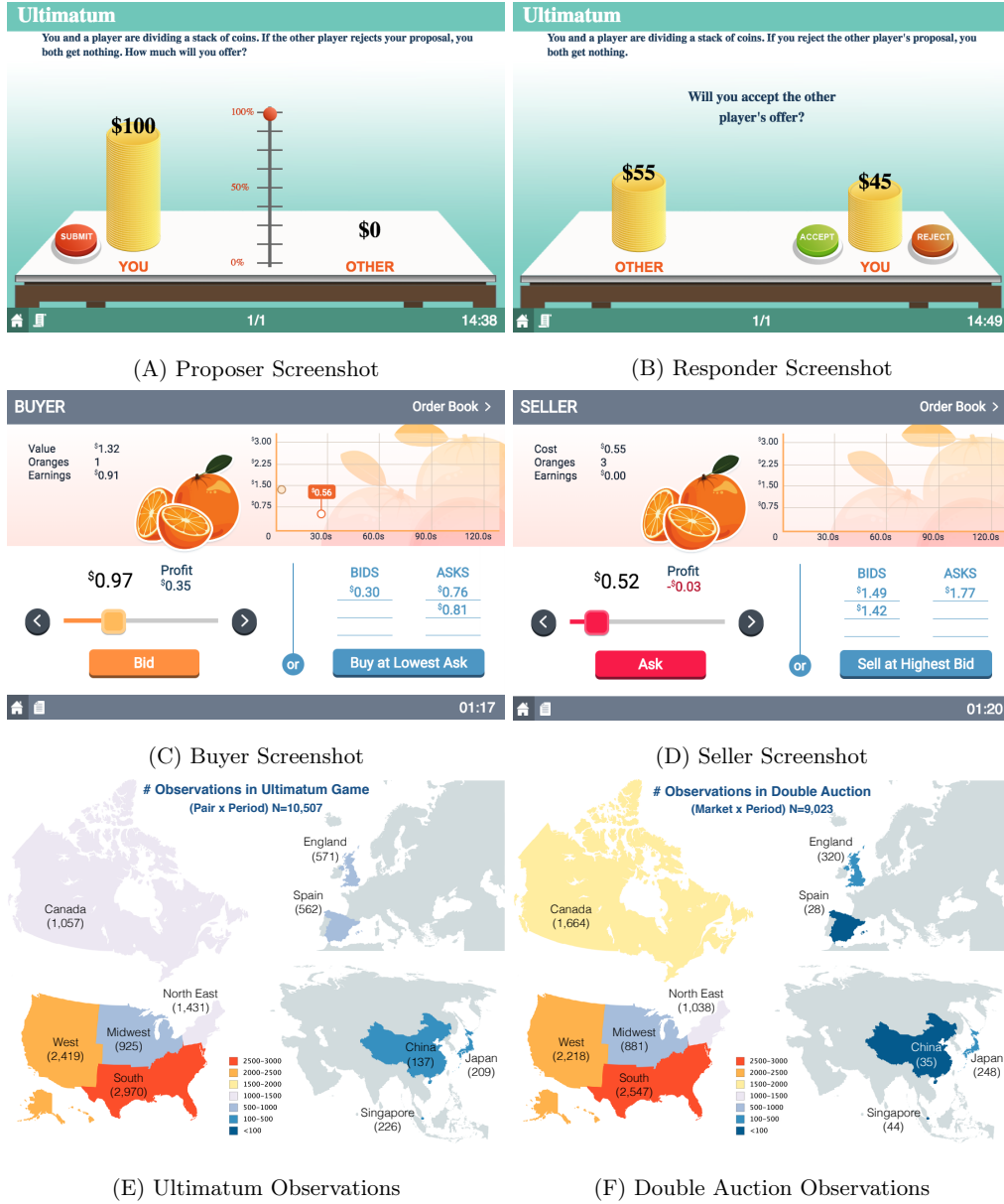


Figure 1: Experimental interface screenshots and geographical distribution of observations. (A and B) The Proposer's and the responder's screen in the ultimatum game. (C and D) The buyer and the seller screens in the buyer-seller centralized trading markets. (E and F) The geographical distribution of observations in ultimatum bargaining game and buyer-seller trading market sessions. The number of observations in each country is in parentheses.

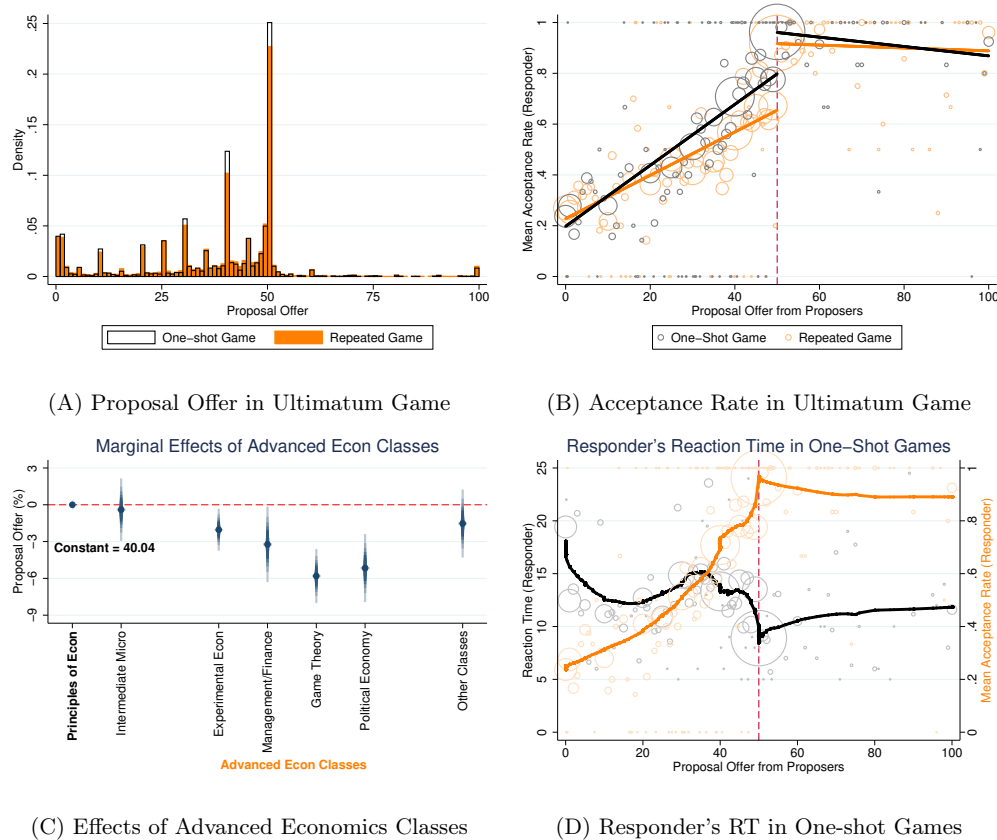
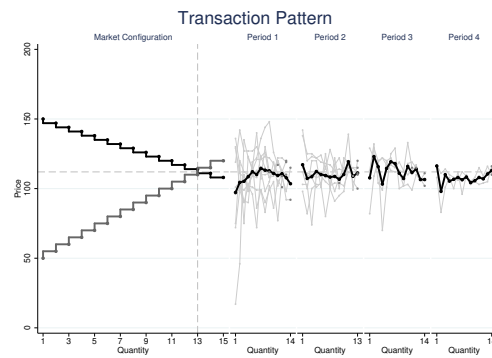
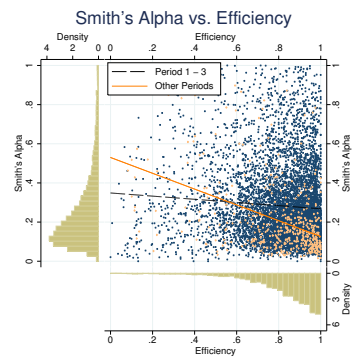


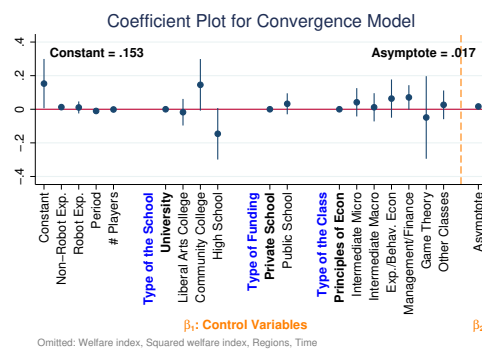
Figure 2: Offers and acceptance rates in ultimatum game experiment. (A) The histogram of proposal offers in one-shot games and repeated games (with a fixed partner protocol). (B) The average acceptance rate for different proposal offer levels, and a piecewise linear regression allowing a discontinuous change at 50%. The size of the dots represents the frequency of each offer amount. (C) The point estimates and 95% CIs of estimated marginal effects of different classes on proposal offers. The principles of economics is the reference group. (D) The LOWESS estimation of the Responder's RT in one-shot games.



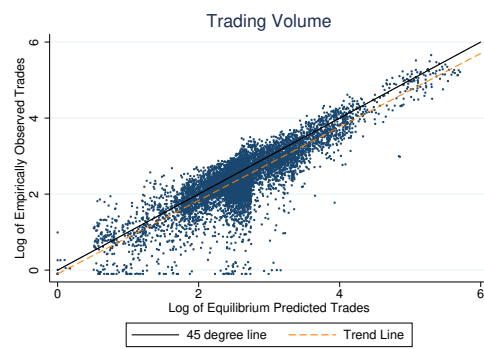
(A) Price Change Visualization



(B) Volatility vs. Efficiency



(C) Convergence of Price



(D) Empirical and Equilibrium Quantity

Figure 3: Design and results of buyer-seller double auctions. (A) There are 5 buyers and 5 sellers who can each trade up to 3 units. The demand curve is the step function plotting the total number of units buyers can afford to buy at different prices (darker line). It starts with one unit of demand at the price of 150, and steps down to 15 units of demand at 108. The supply curve goes up from a single unit of profitable supply at a price of 50, to 15 units of supply at 120 (lighter line). The CE price range is $[110, 114]$ with a midpoint of 112 and generates 13 units of trade. In periods 1-4, each gray line plots real-time averaged time series data of trade prices for one of the 10 different countries for the pictured default market configuration. The number of grey lines are 10, 8, 5, and 2, since most classes play only one or two periods. (B) Joint and marginal distributions of Smith's alpha and efficiency. Regression lines are shown for periods 1 to 3 (black dashed line) and for periods 4 and beyond (orange line). (C) The point estimates and 95% CIs of estimated deviations in initial prices from the convergence model (where the standardized price difference is the dependent variable). For example, in the community college sample prices tend to start higher and in the high school sample prices tend to start lower. (D) A scatter plot of equilibrium and empirical trading volume. The horizontal and vertical axis are the log of equilibrium and empirical quantity, respectively. The black line is the 45 degree line and the orange dotted line is the linear regression line ($R^2 = 0.760$, $\hat{\beta} = 0.967$ and clustered S.E. = 0.012).

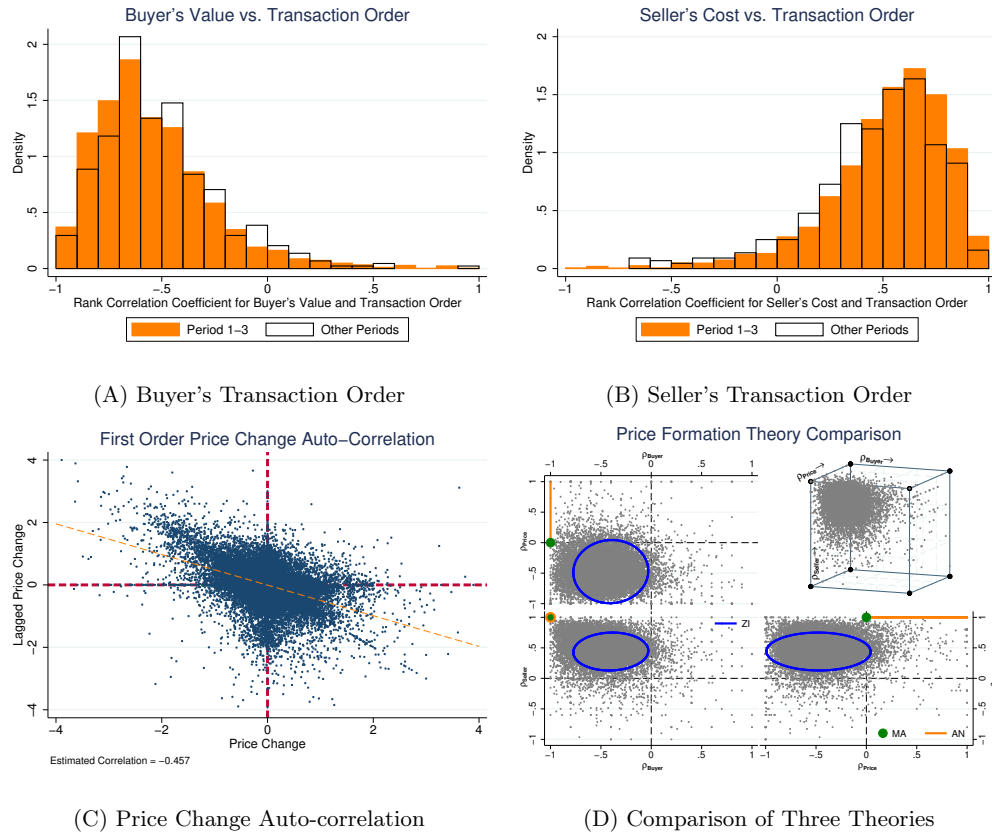


Figure 4: Statistical analysis of trading dynamics. Two statistics can be used to test three hypotheses about trading dynamics: Rank-order correlation and price change auto-correlation. (A) The distribution of Spearman's rank correlation coefficient between buyer's value rank and transaction order. Each observation is one trading period: Orange for period 1 to 3 ($N_{1-3} = 8,052$) and hollow for period 4 and beyond ($N_{\text{other}} = 440$). The average rank-order correlation coefficient is $\rho_{\text{Buyer}} = -0.539$ with p-value < 0.001 ($\rho_{\text{Buyer}}^{1-3} = -0.540$ and $\rho_{\text{Buyer}}^{\text{other}} = -0.506$). (B) The distribution of Spearman's rank correlation coefficient between seller's cost rank and transaction order. (orange: $N_{1-3} = 8,052$, hollow: $N_{\text{other}} = 440$). The average rank-order correlation coefficient is $\rho_{\text{Seller}} = 0.518$ with p-value < 0.001 ($\rho_{\text{Seller}}^{1-3} = 0.521$ and $\rho_{\text{Seller}}^{\text{other}} = 0.472$). (C) The scatter plot of the trade-to-trade price change within a period, correlated with the lagged price change. The overall trade-to-trade price change auto-correlation $\rho_{\text{Price}} = -0.457$; the linear regression line has a slope of -0.491 . (D) The 3-dimensional scatter plot and three-way view of ρ_{Buyer} , ρ_{Seller} and market-level price change auto-correlation ρ_{Price} (with an average of -0.431). Theoretical predictions are depicted in the 3-dimensional scatter plot for Mutual Adjustment (MA, green), Against Nature (AN, orange) and Zero-Intelligence (ZI, blue). The figure on the top-left is the view from above. The figure on the bottom-left is the perspective from the front, and the figure on the bottom right shows what you will see from the right.

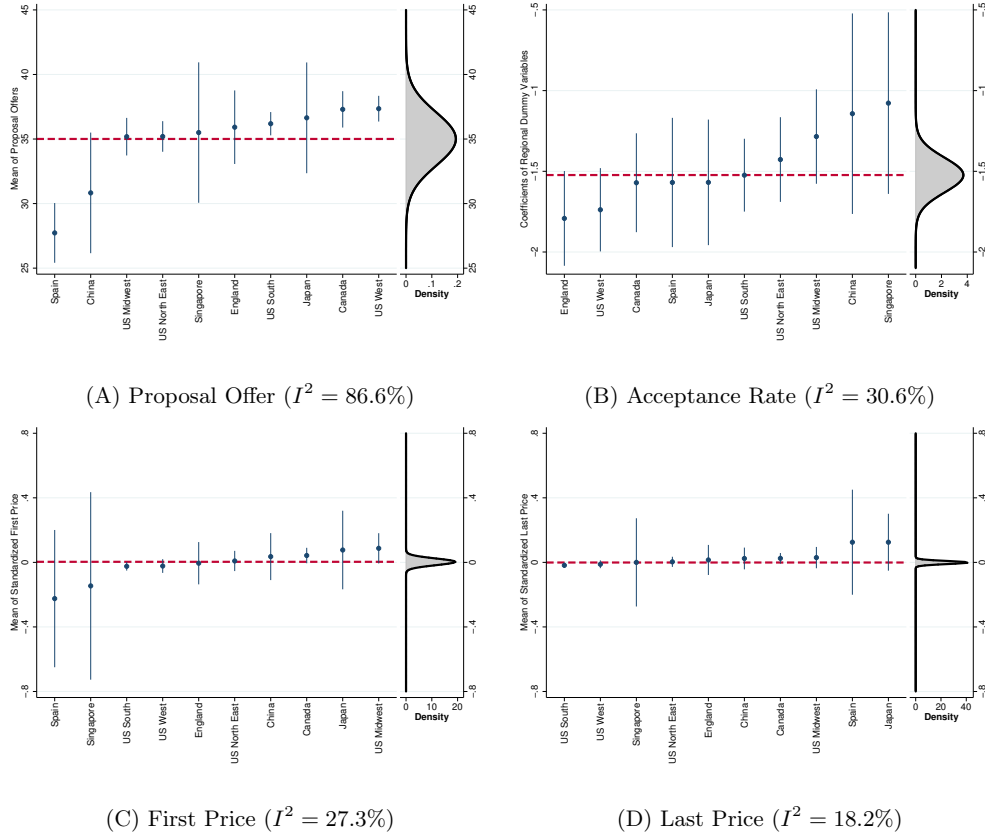


Figure 5: Cross-country heterogeneity in economic bargaining and market behavior. (A) A plot of the means and 95% CIs of the proposal offer deviations for different countries. (B) A plot of the means and 95% CIs of difference in acceptance rates (ultimatum game logit regression intercepts, controlling for offers). (C and D) Plots of the means and 95% CIs of the first and last prices (standardized by equilibrium price) in different countries. I^2 for each plot are reported in parentheses. The distribution of the true means is displayed on the right hand side of each graph.

Online Appendix for General Economic Principles of Bargaining and Trade: Evidence from 2,000 Classroom Experiments

A Ultimatum Game

This section provides the details of ultimatum game analysis. First, we report the detection of the discontinuous jump at equal-split offer. In addition to piecewise linear regressions, we also add quadratic terms to test for robustness. Secondly, we compare our data with RPOZ by conducting the analysis on both data sets. In the third section, we report the OLS regression results for proposal offers with controlling different scenarios and demographic variables. Lastly, we analyze the player's reaction time, finding a drastic drop in the time Responders take to make their decisions when they face the equal-split offer, i.e., 50 out of 100. Table S1 provides descriptive statistics for our ultimatum game data. We summarize the configurations (pie size and number of periods) and players' performance in different subgroups.

Table S1: Descriptive Statistics for Ultimatum Game Data

| Game Configurations | # Pairs | Mean | S.D. | Min | Max |
|-----------------------|---------|----------------|-------|-----------------|-------|
| Pie Size | 6505 | 111.6 | 293.0 | 5 | 5000 |
| Total periods | 6505 | 1.615 | 1.810 | 1 | 20 |
| Performance | N | Proposal Offer | S.D. | Acceptance Rate | S.D. |
| Overall Performance | 10507 | 36.82 | 18.16 | 0.640 | 0.480 |
| Country | | | | | |
| US North East | 1431 | 36.32 | 18.16 | 0.628 | 0.484 |
| US Midwest | 925 | 36.97 | 16.78 | 0.685 | 0.465 |
| US South | 2970 | 36.93 | 18.80 | 0.649 | 0.477 |
| US West | 2419 | 38.02 | 17.43 | 0.647 | 0.478 |
| Canada | 1057 | 37.58 | 19.65 | 0.667 | 0.472 |
| Spain | 562 | 31.19 | 15.34 | 0.532 | 0.499 |
| England | 571 | 37.81 | 16.63 | 0.576 | 0.495 |
| Singapore | 226 | 34.70 | 21.96 | 0.628 | 0.484 |
| Japan | 209 | 38.82 | 18.65 | 0.660 | 0.475 |
| China | 137 | 31.38 | 17.44 | 0.635 | 0.483 |
| Type of School | | | | | |
| University | 9362 | 36.64 | 18.37 | 0.633 | 0.482 |
| Liberal Arts College | 688 | 36.44 | 17.73 | 0.688 | 0.464 |
| Community College | 207 | 42.99 | 12.38 | 0.739 | 0.440 |

Table S1 continued from previous page

| | | | | | |
|------------------------|------|-------|--------|-------|-------|
| High School | 250 | 39.52 | 13.82 | 0.708 | 0.456 |
| Type of Funding | | | | | |
| Private School | 2986 | 36.61 | 18.12 | 0.659 | 0.474 |
| Public School | 7521 | 36.91 | 18.18 | 0.632 | 0.482 |
| Type of Class | | | | | |
| Principles of Econ | 2441 | 38.91 | 16.84 | 0.637 | 0.481 |
| Intermediate Micro | 1110 | 38.58 | 18.20 | 0.694 | 0.461 |
| Exp. / Behav. Econ | 2692 | 37.77 | 19.24 | 0.664 | 0.472 |
| Management / Finance | 938 | 35.88 | 17.93 | 0.630 | 0.483 |
| Game Theory | 2600 | 33.31 | 178.24 | 0.596 | 0.491 |
| Political Economy | 162 | 34.67 | 16.23 | 0.605 | 0.490 |
| Other Classes | 564 | 38.19 | 16.31 | 0.663 | 0.473 |
| Time | | | | | |
| Morning | 4332 | 37.04 | 17.29 | 0.653 | 0.476 |
| Afternoon | 6175 | 36.67 | 18.73 | 0.631 | 0.483 |

A.1 The Detection of Discontinuous Jump

In this subsection, we provide the estimation results of the discontinuous jump. In order to check whether there is a discontinuous jump at the 50% offer, we generate the dummy variable “Fifty or Greater” which is one when the proposal offer is greater or equal to 50. Moreover, to check the behavioral difference in one-shot games and repeated games, we generate another dummy variable “one-shot” which is one when this is an one-shot scenario. The piecewise linear regression results are summarized in Table S2.

Table S2: The Estimation Results of the Discontinuous Jump

| Acceptance Rate | (1) | (2) | (3) | (4) |
|--------------------------|------------------|------------------|----------------------|----------------------|
| Constant | 0.213 [0.018] | 0.228 [0.030] | 0.258 [0.018] | 0.272 [0.030] |
| Offer | 0.010 [0.001] | 0.009 [0.001] | 0.0035 [0.0017] | 0.002 [0.003] |
| Offer ² | | | 0.00014 [0.00003] | 0.00013 [0.00006] |
| Fifty or Greater | 0.791 [0.034] | 0.717 [0.050] | 1.224 [0.227] | 1.087 [0.309] |
| Offer × Fifty or Greater | -0.012 | -0.009 | -0.019 | -0.015 |

Table S2 continued from previous page

| | [0.001] | [0.001] | [0.007] | [0.010] |
|---|---------|---------|-----------|-----------|
| Offer ² ×Fifty or Greater | | | -0.00004 | -0.00004 |
| | | | [0.00006] | [0.00009] |
| One-Shot | | -0.030 | | -0.027 |
| | | [0.034] | | [0.035] |
| One-Shot×Offer | | 0.003 | | 0.002 |
| | | [0.001] | | [0.003] |
| One-Shot×Offer ² | | | | 0.00002 |
| | | | | [0.00007] |
| One-Shot×Fifty or Greater | | 0.138 | | 0.178 |
| | | [0.068] | | [0.439] |
| One-Shot×Offer×Fifty or Greater | | -0.005 | | -0.005 |
| | | [0.002] | | [0.014] |
| One-Shot×Offer ² ×Fifty or Greater | | | | -0.00001 |
| | | | | [0.00012] |
| Discontinuous Jump (Overall) | 0.214 | | 0.166 | |
| | [0.015] | | [0.019] | |
| Discontinuous Jump (One-Shot) | | 0.163 | | 0.106 |
| | | [0.015] | | [0.019] |
| Discontinuous Jump (Repeated) | | 0.262 | | 0.218 |
| | | [0.028] | | [0.034] |
| R-Squared | 0.2344 | 0.2415 | 0.2373 | 0.2445 |
| Observations | 10507 | 10507 | 10507 | 10507 |

The standard errors (shown in parentheses) are clustered at the session level (# of clusters is 490). Moreover, we adopt Wald test for the significance of the discontinuous jump. Estimation results are visualized in Figure S1.

From the estimation of the first two specifications, we can see that the size of the discontinuous jump (see Figure S1A) on average is about 21.4% ($p < 0.001$). When we consider data from one-shot and repeated games separately, we observe that the discontinuous jump in repeated games and one-shot games (see Figure S1B) are 26.2% and 16.3%, respectively (both $p < 0.001$).

We also adopt a quadratic specification (shown in the last two columns). The specification estimates a 16.6% discontinuous jump (see Figure S1C). Estimating the discontinuous jump in one-shot games and repeated games separately, we note a 10.6% jump in one-shot games, and a 21.8% jump in repeated games (see Figure S1D).

Instead of a quadratic specification, one could also further separate out offers of 50% and run a three-part linear regression. We obtain nearly identical discontinuous jumps immediately

before 50%, but insignificant drops immediately after. In particular, there is a discontinuous jump of 21.7% ($p < 0.001$) at 50%, but an insignificant drop of 2.68% ($p = 0.171$) after 50% (see Figure S1E). Consider the data separately, we obtain discontinuous jumps of 16.5% ($p < 0.001$) for one-shot games and 26.5% ($p < 0.001$) for repeated games at 50%, and insignificant drops of 2.57% ($p = 0.188$) for one-shot games and 1.78% ($p = 0.557$) for repeated games after 50% (see Figure S1F).

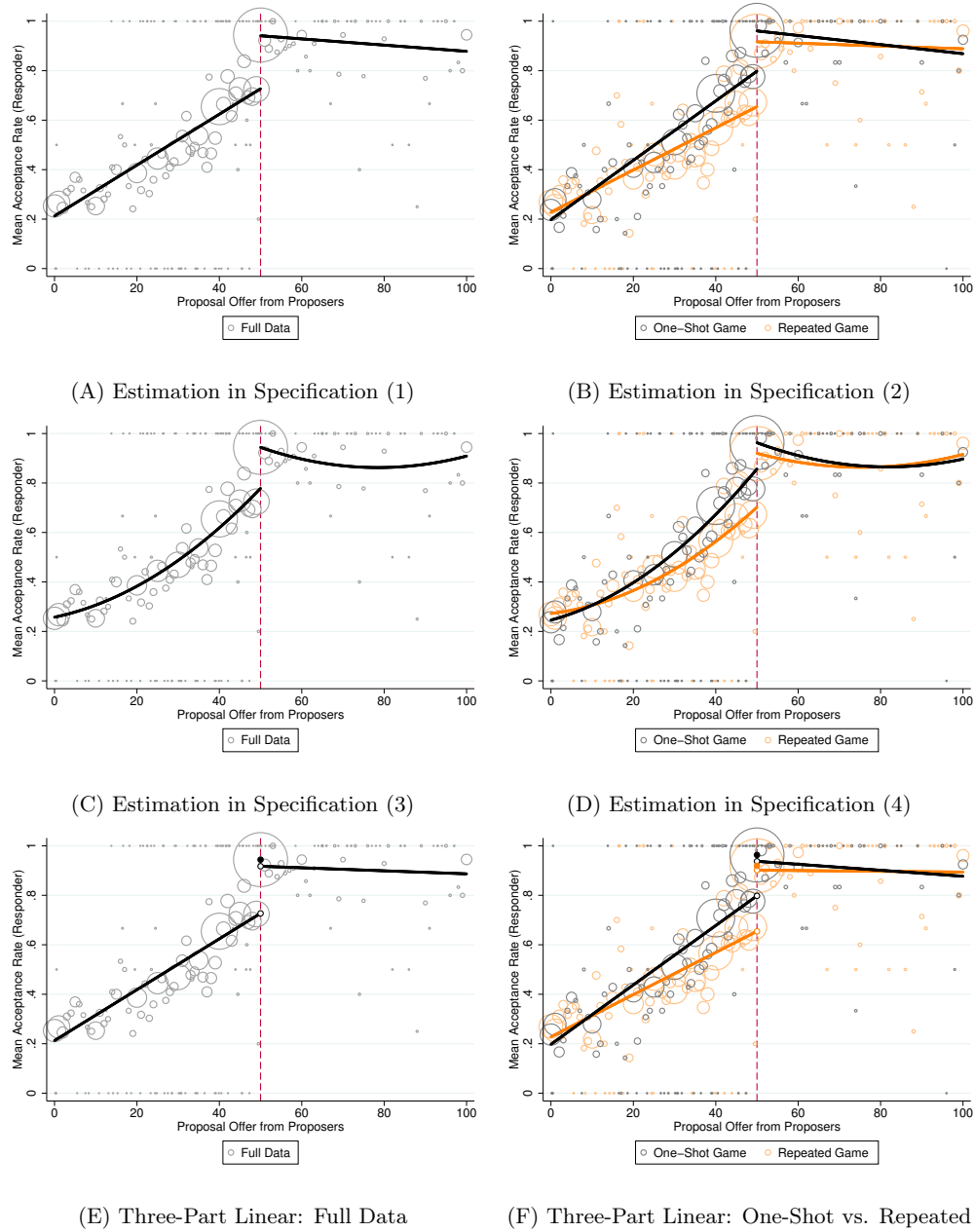


Figure S1: Visualization of the discontinuous jumps under different specifications.

A.2 Comparison between Our Data and RPOZ

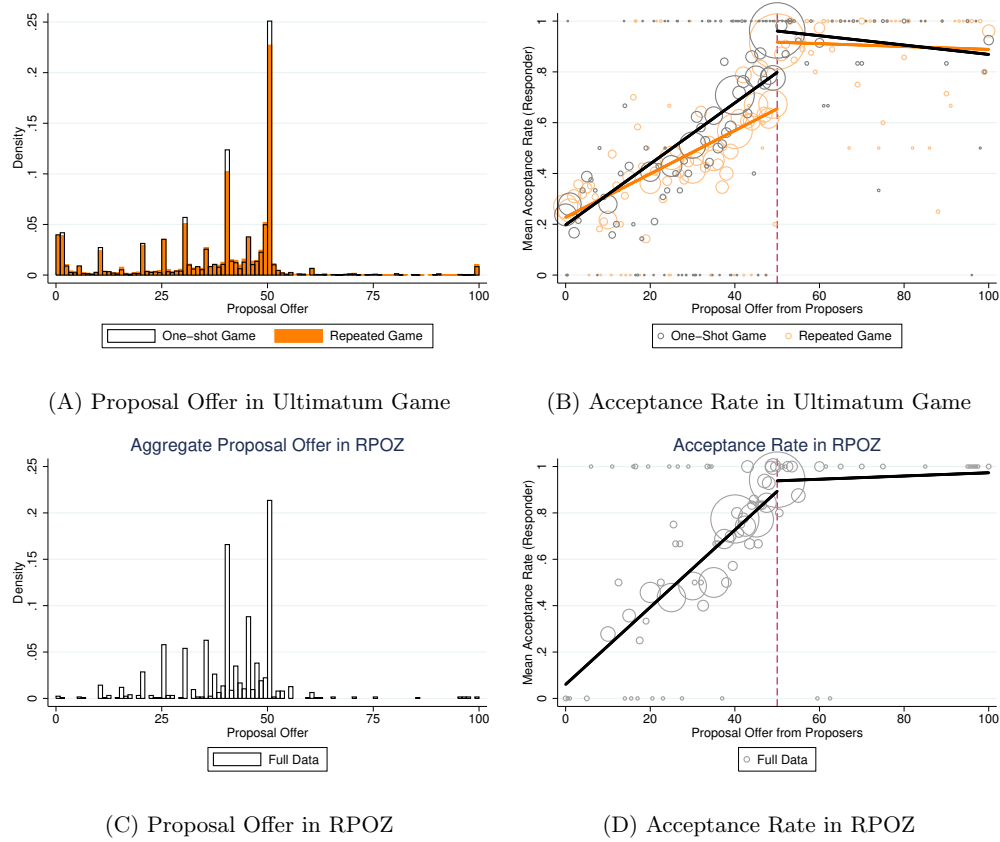


Figure S2: Proposal offers and acceptance rates in our data and RPOZ. (A) The histogram of proposal offers in one-shot games and repeated games (with a fixed partner protocol). (B) The average acceptance rate for different proposal offer levels, and a piecewise linear regression allowing a discontinuous change at 50%. The size of the dots represents the frequency of each offer amount. (C and D) The histogram of proposal offers in RPOZ, and their average acceptance rates.

In this section, we compare our results with RPOZ by conducting the same analysis. The results of RPOZ are reproduced in Figure S2C and S2D. General patterns of both offer amounts and acceptance rates of our data are quite similar to RPOZ except for low offers (which are made and accepted more often) and the detection of the sharp discontinuity. Approximately 8.5% (890 out of 10,507 observations) of the offers are at the equilibrium prediction of the smallest amount or zero, compared to only 0.3% (four out of 1,260 observations) in RPOZ. Moreover, the acceptance rate for these equilibrium offers is 26%. In RPOZ, the discontinuous jump is only 4.5% which is not significant.

A.3 Proposal Offer Regression Analysis

In this subsection, we report OLS regression results for proposal offers (in percentages). Besides geo-demographic dummy variables, we also control for different scenarios such as one-shot games and different rounds in repeated games. Moreover, since Proposers are only allowed to make proposal offers in integers, they are forced to round up their proposal offers when the pie size is small. Therefore, we include a dummy variable for small pie size (≤ 20) to control the effect of a more discrete choice set.

Table S3: Proposal Offer Regression Table

| Proposal Offer (%) | Coef. | Std. Err. | [95% Conf. Interval] | |
|---------------------------|--------------|------------------------|-----------------------------|--------|
| Constant | 40.04 | [1.285] | 37.51 | 42.56 |
| One-Shot | -0.884 | [0.718] | -2.295 | 0.527 |
| First Round Repeated Game | -1.031 | [0.799] | -2.601 | 0.538 |
| Last Round Repeated Game | -0.757 | [0.702] | -2.137 | 0.622 |
| Small Pie | 4.273 | [1.051] | 2.208 | 6.339 |
| US North East | | Reference Group | | |
| US Midwest | -0.223 | [1.333] | -2.842 | 2.395 |
| US South | -1.604 | [1.068] | -3.702 | 0.494 |
| US West | -0.707 | [1.004] | -2.680 | 1.267 |
| Canada | 0.194 | [1.173] | -2.111 | 2.499 |
| Spain | -3.151 | [2.355] | -7.778 | 1.477 |
| England | -1.400 | [2.411] | -6.138 | 3.338 |
| Singapore | -3.182 | [2.429] | -7.955 | 1.590 |
| Japan | 0.711 | [1.857] | -2.937 | 4.359 |
| China | -5.810 | [2.691] | -11.10 | -0.522 |
| University | | Reference Group | | |
| Liberal Arts College | -0.697 | [1.170] | -2.996 | 1.602 |
| Community College | 2.794 | [3.591] | -4.262 | 9.851 |
| High School | 2.161 | [2.607] | -2.961 | 7.283 |
| Private School | | Reference Group | | |
| Public School | 0.852 | [0.900] | -0.916 | 2.620 |
| Principles of Econ | | Reference Group | | |
| Intermediate Micro | -0.399 | [1.296] | -2.944 | 2.147 |
| Exp. / Behavioral Econ | -2.035 | [0.869] | -3.742 | -0.327 |
| Management / Finance | -3.227 | [1.565] | -6.303 | -0.152 |
| Game Theory | -5.806 | [1.122] | -8.011 | -3.602 |

Table S3 continued from previous page

| | | | | |
|-------------------|------------------------|---------|--------|--------|
| Political Economy | -5.163 | [1.411] | -7.935 | -2.392 |
| Other Classes | -1.528 | [1.411] | -4.300 | 1.243 |
| <hr/> | | | | |
| Morning | Reference Group | | | |
| Afternoon | -0.552 | [0.713] | -1.953 | 0.850 |
| <hr/> | | | | |
| Observations | 10,507 | | | |
| R-Squared | 0.0274 | | | |
| <hr/> | | | | |

¹ Small Pie is a dummy which equals to 1 if the pie size is less or equal to 20.

² The standard errors are clustered at the session level (# of clusters is 490).

A.4 Reaction Time Analysis

In this subsection, we analyze the Responders and Proposers' RT. To control the effect of experience, we separate the data into four different scenarios: "one-shot game," "repeated game first round," "repeated game non-last round" and "repeated game last round." After all, once players play the game repeatedly, they will be more familiar with the game and take shorter time to react.

First, we analyze Responders' reaction time when they are facing different proposal offers from the Proposer. To ensure our analysis is not skewed by extreme values, we drop 15 outliers (out of 10,507 observations) that take more than 180 seconds to respond. In Figure S3, we plot the LOWESS estimation of Responders' reaction time for different proposal offers. In the one-shot game scenario, we find an inverted-U shape for RT when the offer increases from 20% to 50% where acceptance rate increases from 40% to nearly 100%. This population-level result is consistent with the individual results of Krajbich et al. (2014) that find 16 out of 18 subjects have their average RT peak at the offer they are indifferent when eliciting acceptance decision 3–5 times for four different offers, 20%, 30%, 40%, and 50%. However, RT increases are evident in all scenarios as the offer drops below 20%, even though acceptance rates plummet. Finally, we observe there is a sharp drop of reaction time at exactly 50% in all scenarios. For instance, in the one-shot game scenario, Responders' average reaction time is about 13.5 seconds when they encounter a 49% offer. Yet, it only takes them 8.9 seconds to respond when they are facing an equal-split offer.

Secondly, we turn to analyze Proposers' reaction time. Proposers make their decisions based on the expected acceptance rate of different proposal offers. Therefore, here we analyze the relationship between Proposers' reaction time and the expected payoff. To ensure our analysis not to be skewed by outliers, we exclude 40 choices (out of 10507 observations) that take more than 180 seconds to make the decision. Our results are plotted in Figure S4. Intuitively, the higher expected payoff means the proposal offer is more attractive to the Proposer. Therefore, the Proposer should take shorter reaction time to make the decision. From the graph, we can observe that this decreasing pattern in one-shot games and the first round of repeated games. Yet, this trend is not obvious in other periods of repeated games.

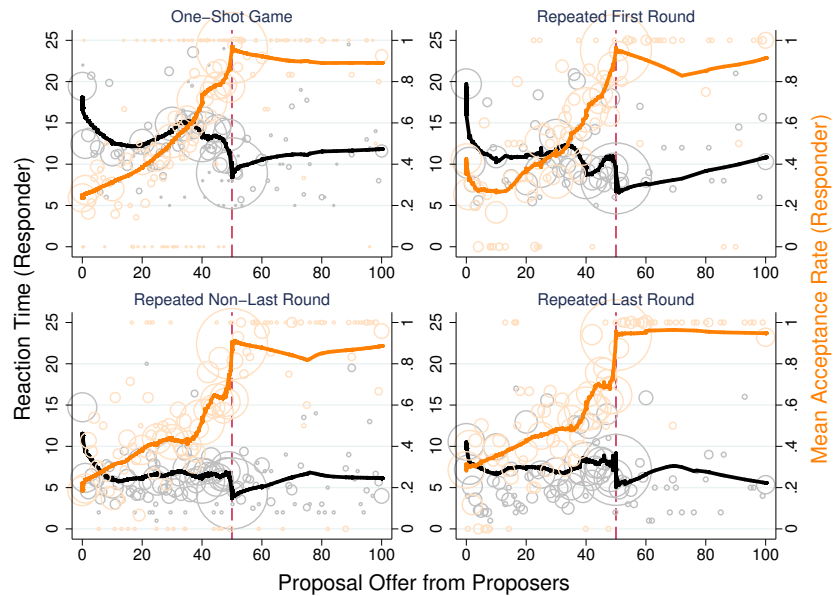


Figure S3: The LOWESS estimation (bandwidth equals to 0.2) of Responders' acceptance rate (shown in orange) and reaction time (shown in black). The acceptance rate is shown by the vertical axis on the right and the Responders' reaction time is presented by the vertical axis on the left. The size of the dots represents the frequency of that proposal offer.

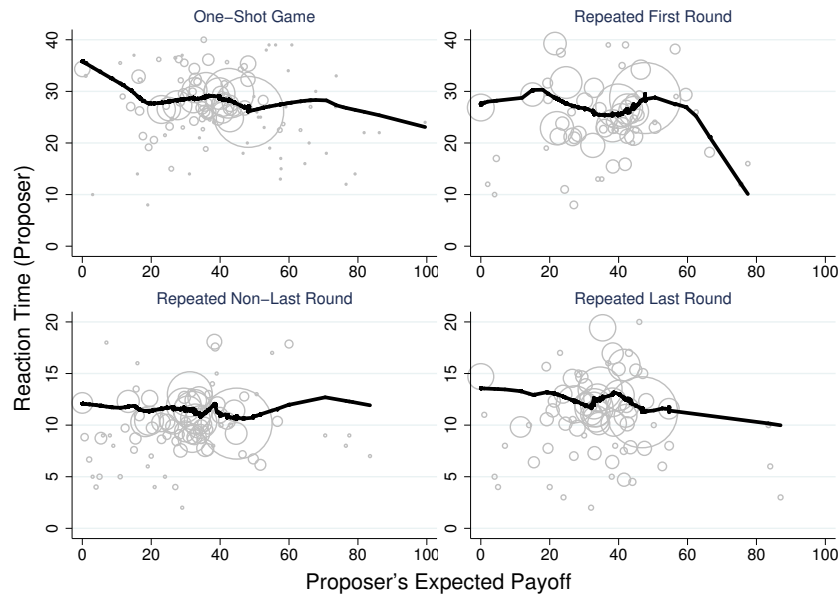


Figure S4: The LOWESS estimation (bandwidth equals to 0.5) of Proposers' expected payoff and reaction time. The size of the dots is weighted by the frequency of the expected payoff.

B Double Auction

In this section, we provide details of our analysis for buyer-seller double auction. We first report the full regression table of within-period price convergence. Furthermore, we analyze how the market would converge if the market were played multiple periods by estimating the exponential convergence model. Lastly, we describe the theories and predictions of price dynamics and corresponding descriptive statistics of relevant correlation coefficients.

One market session was excluded due to an average efficiency level of -4092.0% caused by an input error when demonstrating the effect of a supply shift using the fairly normal supply-demand schedule of Figure S5A. By mistyping 136 as 1136 when entering maximum seller cost, the instructor created a 50-player market with an equilibrium quantity of merely 2 units (Figure S5B). This extraordinary setting resulted in an average of 0.08 equilibrium trades per player, much lower than all other markets (with at least 0.4 trades per player). With little or no financial consequences for trading at a loss, excessive extra-marginal trades occurred that lead to negative efficiency levels in all of the nine periods the market was repeated, ranging from -1685.1% to -6219.1% .

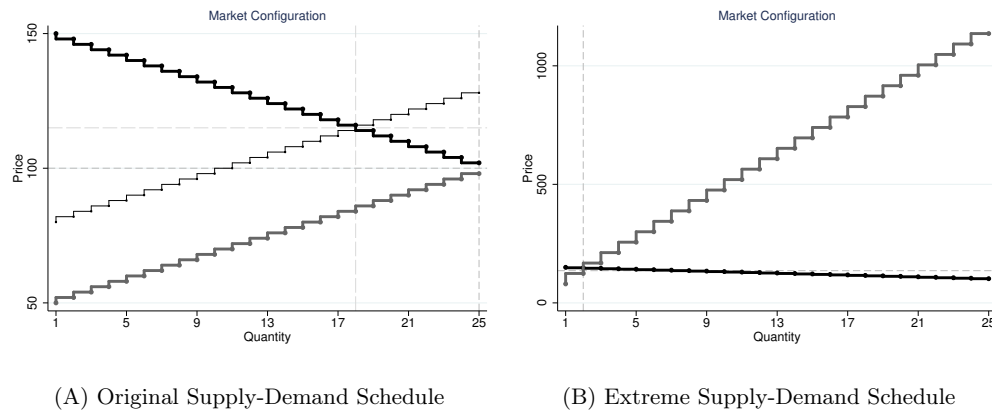


Figure S5: (a) The intended supply-demand schedule to illustrate the effect of a supply shift. (b) The extreme schedule from a supply shift where 1136 (instead of 136) was entered as maximum seller cost. This resulted in numerous dominated trades and negative efficiency levels, something rarely seen when subjects are incentivized with financial rewards.

In Table S4, we summarize our double auction data. In the first panel, we provide aggregate market information. These information contains the experience of the players in the market and the configurations. In the second panel, we provide descriptive statistics market performance, including mean error deviation ($\bar{\delta}$), Smith's alpha (α), and efficiency (E) in different subgroups.

Table S4: Descriptive Statistics for Double Auction Data

| Market Information | # Markets | | Mean | S.D. | Median | Min | Max |
|----------------------|-----------|----------------|-------|----------|--------|-------|-------|
| # of Players | 5809 | | 14.11 | 15.43 | 10 | 2 | 318 |
| Non-Robot Experience | 5809 | | 2.529 | 2.781 | 1.7 | 0 | 32.33 |
| Robot Experience | 5809 | | 0.245 | 0.634 | 0 | 0 | 11.67 |
| Total Periods | 5809 | | 1.553 | 0.959 | 1 | 1 | 10 |
| Welfare Index | 5809 | | 0.188 | 0.111 | 0.231 | 0 | 0.901 |
| Market Performance | N | $\bar{\delta}$ | S.D. | α | S.D. | E | S.D. |
| Overall Performance | 9023 | 0.070 | 0.280 | 0.279 | 0.294 | 0.815 | 0.258 |
| Country | | | | | | | |
| US North East | 1038 | 0.044 | 0.203 | 0.234 | 0.204 | 0.824 | 0.167 |
| US Midwest | 881 | 0.063 | 0.244 | 0.290 | 0.286 | 0.803 | 0.230 |
| US South | 2547 | 0.041 | 0.301 | 0.248 | 0.336 | 0.811 | 0.385 |
| US West | 2218 | 0.021 | 0.200 | 0.264 | 0.204 | 0.822 | 0.196 |
| Canada | 1664 | 0.209 | 0.377 | 0.393 | 0.374 | 0.836 | 0.143 |
| Spain | 28 | -0.058 | 0.258 | 0.274 | 0.199 | 0.860 | 0.228 |
| England | 320 | 0.041 | 0.130 | 0.210 | 0.162 | 0.727 | 0.156 |
| Singapore | 44 | -0.005 | 0.128 | 0.201 | 0.139 | 0.780 | 0.166 |
| Japan | 248 | 0.055 | 0.200 | 0.238 | 0.208 | 0.753 | 0.238 |
| China | 35 | 0.066 | 0.098 | 0.136 | 0.068 | 0.848 | 0.163 |
| Type of School | | | | | | | |
| University | 8368 | 0.074 | 0.285 | 0.279 | 0.299 | 0.812 | 0.264 |
| Liberal Arts College | 523 | 0.010 | 0.161 | 0.243 | 0.178 | 0.864 | 0.149 |
| Community College | 65 | 0.069 | 0.426 | 0.534 | 0.353 | 0.695 | 0.259 |
| High School | 67 | -0.063 | 0.184 | 0.272 | 0.196 | 0.885 | 0.135 |
| Type of Funding | | | | | | | |
| Private School | 1769 | 0.026 | 0.196 | 0.233 | 0.213 | 0.810 | 0.450 |
| Public School | 7254 | 0.080 | 0.296 | 0.290 | 0.309 | 0.816 | 0.183 |
| Type of Class | | | | | | | |
| Principles of Econ | 6490 | 0.074 | 0.262 | 0.279 | 0.266 | 0.816 | 0.273 |
| Intermediate Micro | 1045 | 0.080 | 0.426 | 0.314 | 0.472 | 0.814 | 0.242 |
| Intermediate Macro | 249 | 0.024 | 0.247 | 0.268 | 0.253 | 0.817 | 0.187 |
| Exp. / Behav. Econ | 437 | 0.037 | 0.187 | 0.218 | 0.197 | 0.800 | 0.210 |
| Management / Finance | 452 | 0.051 | 0.234 | 0.263 | 0.265 | 0.791 | 0.216 |
| Game Theory | 20 | -0.057 | 0.226 | 0.295 | 0.175 | 0.814 | 0.216 |

Table S4 continued from previous page

| | | | | | | | |
|-------------------|------|-------|-------|-------|-------|-------|-------|
| Political Economy | 2 | 0.766 | 0.701 | 1.030 | 0.433 | 0.605 | 0.047 |
| Other Classes | 328 | 0.048 | 0.214 | 0.272 | 0.220 | 0.838 | 0.144 |
| <hr/> | | | | | | | |
| Time | | | | | | | |
| Morning | 783 | 0.072 | 0.330 | 0.321 | 0.361 | 0.789 | 0.334 |
| Afternoon | 8240 | 0.069 | 0.275 | 0.275 | 0.286 | 0.817 | 0.250 |

¹ Non-Robot Experience: The average number of trading periods that the players played *with only humans* before this market starts.

² Robot Experience: The average number of periods that the players played in the market *including robot players* before this market starts.

³ Welfare Index: $w = |\Pi_c - \Pi_p| / (\Pi_c + \Pi_p)$, where Π_c and Π_p are the total consumer and producer surplus. We construct this variable to control the configuration.

⁴ Mean Error Deviation is defined as $\bar{\delta} = \frac{1}{Q} \sum_{q=1}^Q \delta_q$, which is the average standardized price difference.

B.1 Within-Period Convergence

We follow [Noussair et al. \(1995\)](#) to estimate the convergence of price within a transaction period:

$$y_{it} = (1/t)\mathbb{X}_i \cdot \beta_1 + (1 - 1/t)\beta_2 + \epsilon_{it}$$

where i indicates the particular period of market, t represents time as measured by the transaction order, \mathbb{X}_i contains the information about that period of market, and β_1 explains the origin of the dynamic process. Therefore, β_2 is the asymptote of the dependent variable.

We take standardized price difference, which is the difference of transaction price and equilibrium price divided by the equilibrium price, as the dependent variable. Table S5 shows that prices initially are 15.3% higher than equilibrium prices, but eventually drop to only 1.7% higher. This shows strong convergence within the period.

Table S5: Estimation Result of Within-Period Price Convergence

| | Coef. | Std. Err. | [95% Conf. Interval] | |
|---------------------------------|--------|-----------|----------------------|--------|
| Initial Price Variation: | | | | |
| Constant | 0.153 | [0.075] | 0.006 | 0.299 |
| Non-Robot Experience | 0.013 | [0.004] | 0.004 | 0.021 |
| Robot Experience | 0.011 | [0.018] | -0.025 | 0.047 |
| Period | -0.010 | [0.009] | -0.028 | 0.008 |
| Welfare Index | -3.477 | [1.026] | -5.489 | -1.465 |
| Welfare Index Sq. | 11.74 | [3.317] | 5.239 | 18.25 |
| # of Players | -0.001 | [0.001] | -0.003 | 0.001 |

Table S5 continued from previous page

| US North East | | Reference Group | | |
|---------------------------|---------|------------------------|--------|-------|
| US Midwest | 0.100 | [0.054] | -0.007 | 0.206 |
| US South | -0.048 | [0.041] | -0.128 | 0.032 |
| US West | -0.087 | [0.047] | -0.180 | 0.005 |
| Canada | 0.217 | [0.059] | 0.102 | 0.332 |
| Spain | -0.211 | [0.113] | -0.433 | 0.011 |
| England | -0.053 | [0.071] | -0.192 | 0.086 |
| Singapore | -0.114 | [0.272] | -0.647 | 0.419 |
| Japan | -0.044 | [0.074] | -0.188 | 0.100 |
| China | -0.046 | [0.093] | -0.228 | 0.135 |
| University | | Reference Group | | |
| Liberal Arts College | -0.018 | [0.040] | -0.097 | 0.062 |
| Community College | 0.145 | [0.078] | -0.008 | 0.299 |
| High School | -0.146 | [0.078] | -0.299 | 0.007 |
| Private School | | Reference Group | | |
| Public School | 0.033 | [0.032] | -0.030 | 0.095 |
| Principles of Econ | | Reference Group | | |
| Intermediate Micro | 0.042 | [0.043] | -0.043 | 0.126 |
| Intermediate Macro | 0.012 | [0.043] | -0.072 | 0.096 |
| Exp. / Behavioral Econ | 0.064 | [0.058] | -0.050 | 0.177 |
| Management / Finance | 0.071 | [0.037] | -0.002 | 0.143 |
| Game Theory | -0.049 | [0.125] | -0.295 | 0.197 |
| Other Classes | 0.026 | [0.044] | -0.059 | 0.112 |
| Morning | | Reference Group | | |
| Afternoon | 0.030 | [0.041] | -0.051 | 0.111 |
| Ending Point: | | | | |
| Asymptote | 0.017 | [0.006] | 0.006 | 0.028 |
| Observations | 138,898 | | | |
| R-Squared | 0.0966 | | | |

The standard errors (shown in parentheses) are clustered at the session level (# of clusters is 1672). To compute the clustered standard errors, we drop two market observations for Political Economy class since they are from the same market session.

B.2 Between-Period Convergence

We analyze how markets would converge if they were played multiple times. In particular, we follow the approach in Gjerstad (2007) to analyze the convergence of volatility and efficiency using an exponential convergence model. The convergence of the dependent variable is modeled as:

$$y(t) = c \cdot t^{-b} \eta(t)$$

where $y(t)$ is the dependent variable at period t and $\eta(t)$ is the random error assumed to be independent and lognormally distributed. Once we take logarithm on both sides, we can express the exponential model as the following linear model:

$$\ln[y(t)] = \ln(c) - b \cdot \ln(t) + \epsilon(t)$$

where $\epsilon(t)$ is the random error following a normal distribution. Notice that since $y(1) = c$, \hat{c} would be the estimation of starting point and \hat{b} is the estimation of the speed of convergence. The estimation results of Smith's alpha ($\alpha(t)$) and efficiency loss ($1 - E(t)$) are shown in Table S6.

Table S6: Estimation Result of Exponential Convergence Model

| | (1) | (2) | (3) | (4) |
|--------------------|------------------|------------------|-------------------|-------------------|
| Dependent Variable | $\alpha(t)$ | $\alpha(t)$ | $1 - E(t)$ | $1 - E(t)$ |
| \hat{c} | 0.225 [0.004] | 0.206 [0.016] | 0.101 [0.003] | 0.077 [0.012] |
| \hat{b} | 0.292 [0.029] | 0.270 [0.025] | -0.067 [0.053] | -0.012 [0.049] |
| $\hat{\alpha}(25)$ | 0.088 [0.008] | 0.086 [0.009] | | |
| $1 - \hat{E}(25)$ | | | 0.125 [0.021] | 0.081 [0.016] |
| Controls | NO | YES | NO | YES |
| Observations | 9,021 | 9,021 | 9,021 | 9,021 |
| R-squared | 0.032 | 0.129 | 0.001 | 0.042 |

The standard errors (shown in parentheses) are clustered at the session level (# of clusters is 1672). To compute the clustered standard errors, We drop two market observations for Political Economy class since they are from the same market session. The controls contain the experience, number of players in the market, welfare index (and squared), countries, school types, funding types, class types and time.

The result shows that the initial point of Smith's alpha is 20.6% (22.5% without controls) and the efficiency loss is 7.7% (10.1% without controls), which are both significantly different from zero.

The significance of \hat{b} for Smith's alpha shows a convergence trend. However, no convergence trend can be not found for efficiency loss. Based on these results, we can estimate Smith's alpha and efficiency loss if the market were played 25 periods. Price volatility in period 25 is estimated to be 8.6% (8.8% without controls) and efficiency loss is predicted to be 8.1% (12.5% without controls). This suggests that price volatility decreases as the market lasts more periods, but efficiency remains stable across periods.

B.3 Theories of Price Formation

In this section, we describe three theories of within-period price dynamics and provide corresponding theoretical predictions. These theories are the "Mutual Adjustment" (MA) model of [Wilson \(1987\)](#), the "Against Nature" (AN) model of [Friedman \(1991\)](#), and the zero-intelligence algorithm (ZI) proposed by [Gode and Sunder \(1993\)](#). Here we only summarize predictions that we test. A more detailed comparison can be found in [Cason and Friedman \(1996\)](#). For between-period price dynamics, we briefly describe [Easley and Ledyard \(1993\)](#).

Mutual Adjustment (MA)

[Wilson \(1987\)](#) extends the bilateral bargaining model between a buyer and a seller to the context of multilateral markets. Hence, price dynamics in the double auction is viewed as a sequential equilibrium of an extensive-form game where the buyers' values and the sellers' costs are drawn from a commonly known joint distribution. Under this framework, buyers and sellers play a waiting game where players' impatience arises from the pressure of missing a profitable transaction. Eventually, this impatience induces some trader to make a bid or ask that has a positive probability to be accepted in the sequential equilibrium. If the transaction does not happen immediately, the trader would improve the offer until it is accepted. This improvement process is like a Dutch auction. Based on this model, we have the following testable predictions:

1. *Transaction Partner:* The buyers (sellers) with higher values (lower costs) should transact earlier. That is, the rank-order correlation of buyer's value (seller's cost) and transaction order should be negative (positive). Moreover, once we assume risk neutrality and symmetry, the rank-order correlation coefficient ρ_{Buyer} (ρ_{Seller}) should be exactly -1.0 (1.0).
2. *Price Dynamics:* Price changes are serially uncorrelated. Therefore, the first-order price change auto-correlation coefficient ρ_{Price} is predicted to be 0 in this model.

Against Nature (AN)

[Friedman \(1991\)](#) models the price formation process in double auctions as a Bayesian game against nature. This framework aligns with the Bertrand perspective that traders would accept the market offer once it exceeds their own reservation price. Furthermore, traders are assumed to neglect strategic feedback effects—they ignore the impact of their own bids and asks on other players'

strategies. Under this construction, the traders' problem can be transformed into solving the optimal stopping point with respect to Nature's bid and ask generating process. The reservation price solves the problem in equilibrium. This theory implies that the transaction order should be in order of the buyer and seller's value and cost, and price change is positively auto-correlated. These two testable predictions are listed below.

1. *Transaction Partner*: The buyers (sellers) with higher values (lower costs) should transact earlier. That is, the rank-order correlation of buyer's value (seller's cost) and transaction order should be negative (positive). Moreover, if we assume risk neutrality and symmetry, the rank-order correlation coefficient ρ_{Buyer} (ρ_{Seller}) should be exactly -1.0 (1.0).
2. *Price Dynamics*: Price changes are positively auto-correlated. Therefore, the first-order price change auto-correlation coefficient ρ_{Price} is predicted to be positive in this model.

Zero-Intelligence (ZI)

To show how efficient the double auction mechanism is, [Gode and Sunder \(1993\)](#) developed a "zero-intelligence algorithm" where buyers and sellers make random bids and asks (but avoid losing money). Specifically, the buyer's bid is uniformly distributed from 0 to the buyer's value. Similarly, the seller's ask is uniformly distributed from the upper bound to the cost. The construction of ZI algorithm implies that the transaction prices are independently drawn from a distribution that would change as transactions occur. Since bids and asks are completely random in this environment, there is no precise prediction. Yet, we can still infer from the algorithm that it is slightly more likely that buyers with higher values and sellers with lower costs would trade earlier (because they can accept offers from a wider range). Moreover, if we assume that transaction prices come from a fixed distribution, the price change auto-correlation coefficient is exactly -0.5 ¹¹.

Although there is no clear prediction about the rank order correlation coefficients and the price change auto-correlation, we follow [Cason and Friedman \(1996\)](#) and simulate the algorithm in our default setting for 10,000 times. The simulation result is shown in Figure S6 and the blue circles indicate 95% confidence regions. In this simulation data set, the mean of ρ_{price} is -0.475 (S.D. = 0.211). On the other hand, the mean of ρ_{Buyer} and ρ_{Seller} are -0.402 (S.D. = 0.152) and 0.440 (S.D. = 0.127), respectively. Thus, implications of ZI algorithm are:

1. *Transaction Partner*: The buyers with higher values and the sellers with lower costs would be slightly more likely to trade earlier. 10,000 simulations predict $\rho_{\text{Buyer}} = -0.402$ and $\rho_{\text{Seller}} = 0.440$.

¹¹Suppose transaction prices $p_q, q \in \{1, \dots, Q\}$ are IID with variance V . We can normalize p_q such that $\mathbb{E}(p_q) = 0$. Therefore, $V = \mathbb{E}(p_q^2)$ and $\mathbb{E}(p_q p_{q+1}) = \mathbb{E}(p_q) \mathbb{E}(p_{q+1}) = 0$. In this case, the price change auto-correlation coefficient would be

$$\rho_{\text{price}} = \frac{\mathbb{E}[(p_{q+1} - p_q)(p_q - p_{q-1})]}{\mathbb{E}[(p_{q+1} - p_q)^2]} = \frac{-\mathbb{E}(p_q^2)}{\mathbb{E}(p_{q+1}^2) + \mathbb{E}(p_q^2)} = \frac{-V}{2V} = -0.5.$$

After some traders complete the transactions and leave the market, the distribution of prices would change. Yet, the changes in distribution would only have a small effect on ρ_{price} .

2. *Price Dynamics*: The price change auto-correlation coefficient is predicted (by 10,000 simulations) to be $\rho_{\text{Price}} = -0.475$.

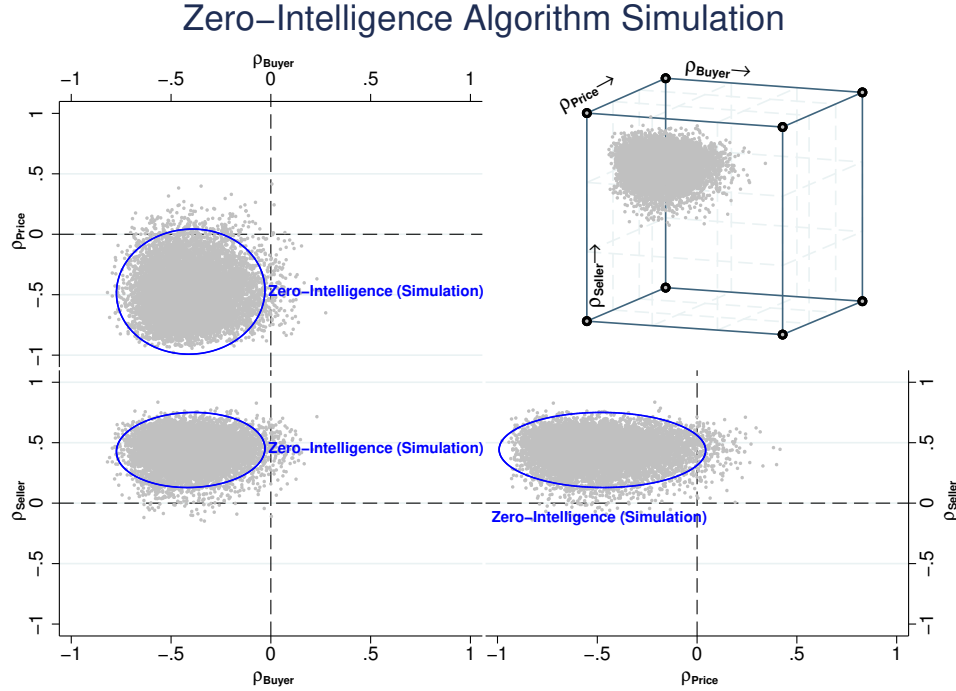


Figure S6: Zero-Intelligence algorithm simulation result.

Between-Period Price Formation

[Easley and Ledyard \(1993\)](#) postulate traders use the price range of last period to form reservation prices they bid up to in a reduced-form English auction in the current period, but then adjust to their true values or costs if they fail to make a trade. This model predicts a sequence of decreasing price ranges across periods, resulting in a distributional shift of rank order correlation between periods. Hence, we have:

1. *Transaction Partner*: The buyers (sellers) with higher values (lower costs) should transact earlier, but this early tendency diminishes across periods.
2. *Price Dynamics*: No prediction for within-period price dynamics (since this is a between-period theory).

B.4 Descriptive Statistics for Price Formation Indices

In Table S7, we provide the descriptive statistics of ρ_{Price} , ρ_{Buyer} and ρ_{Seller} in different groups.

Table S7: Descriptive Statistics for Price Formation Indices

| Correlation Coefficient | N | ρ_{Price} | S.D. | ρ_{Buyer} | S.D. | ρ_{Seller} | S.D. |
|-------------------------|------|-----------------------|-------|-----------------------|-------|------------------------|-------|
| Overall Performance | 8492 | -0.431 | 0.313 | -0.539 | 0.277 | 0.518 | 0.279 |
| Country | | | | | | | |
| US North East | 965 | -0.440 | 0.362 | -0.507 | 0.307 | 0.496 | 0.297 |
| US Midwest | 826 | -0.427 | 0.266 | -0.544 | 0.219 | 0.513 | 0.241 |
| US South | 2418 | -0.425 | 0.326 | -0.552 | 0.278 | 0.541 | 0.279 |
| US West | 2144 | -0.422 | 0.284 | -0.521 | 0.285 | 0.513 | 0.286 |
| Canada | 1620 | -0.444 | 0.306 | -0.572 | 0.254 | 0.519 | 0.267 |
| Spain | 24 | -0.459 | 0.294 | -0.232 | 0.487 | 0.263 | 0.526 |
| England | 311 | -0.454 | 0.251 | -0.547 | 0.212 | 0.498 | 0.214 |
| Singapore | 42 | -0.429 | 0.244 | -0.525 | 0.185 | 0.466 | 0.260 |
| Japan | 108 | -0.428 | 0.595 | -0.370 | 0.480 | 0.493 | 0.411 |
| China | 34 | -0.415 | 0.329 | -0.553 | 0.251 | 0.508 | 0.315 |
| Type of School | | | | | | | |
| University | 7865 | -0.433 | 0.313 | -0.536 | 0.278 | 0.514 | 0.280 |
| Liberal Arts College | 510 | -0.410 | 0.301 | -0.577 | 0.257 | 0.582 | 0.231 |
| Community College | 50 | -0.445 | 0.402 | -0.520 | 0.237 | 0.533 | 0.291 |
| High School | 67 | -0.389 | 0.303 | -0.527 | 0.361 | 0.501 | 0.390 |
| Type of Funding | | | | | | | |
| Private School | 1656 | -0.422 | 0.324 | -0.555 | 0.286 | 0.532 | 0.277 |
| Public School | 6836 | -0.433 | 0.310 | -0.535 | 0.275 | 0.515 | 0.279 |
| Type of Class | | | | | | | |
| Principles of Econ | 6262 | -0.432 | 0.303 | -0.544 | 0.268 | 0.524 | 0.269 |
| Intermediate Micro | 974 | -0.424 | 0.312 | -0.567 | 0.246 | 0.556 | 0.242 |
| Intermediate Macro | 228 | -0.447 | 0.312 | -0.517 | 0.281 | 0.505 | 0.310 |
| Exp. / Behav. Econ | 287 | -0.444 | 0.444 | -0.444 | 0.402 | 0.444 | 0.404 |
| Management / Finance | 422 | -0.417 | 0.330 | -0.521 | 0.302 | 0.471 | 0.307 |
| Game Theory | 19 | -0.402 | 0.280 | -0.597 | 0.163 | 0.443 | 0.289 |
| Political Economy | 1 | -0.275 | – | -0.75 | – | 0.607 | – |
| Other Classes | 299 | -0.415 | 0.343 | -0.463 | 0.336 | 0.430 | 0.345 |
| Time | | | | | | | |
| Morning | 733 | -0.434 | 0.294 | -0.539 | 0.280 | 0.509 | 0.266 |
| Afternoon | 7759 | -0.431 | 0.315 | -0.539 | 0.277 | 0.519 | 0.280 |

C Heterogeneity Analysis

We report details of the cross-country heterogeneity analysis in this section. We analyze the heterogeneity in the first period of the default configuration for both ultimatum game and double auction to control for experience and experimental setup. In the ultimatum game, we analyze the heterogeneity in proposal offers and acceptance rates. On the other hand, in double auctions, we study the heterogeneity of the first and the last price in a market period.

We apply the DerSimonian-Laird random-effects model of meta-analysis, and calculate the I^2 statistic, which shows the proportion of the variability from between-country heterogeneity. Table S8, S10, S12 and S14 report the means and 95% CIs in each country for proposal offer, acceptance rate, first transaction price and last transaction price, respectively. Table S9, S11, S13 and S15 report (for the same four indices) measures of heterogeneity, including the I^2 statistic, the estimated variance of the true means $\hat{\tau}^2$, as well as the p -value for the null hypothesis of having no between-country heterogeneity.

Specifically, under the framework of random-effect model, the mean of country k is

$$m_k = (\mu + \zeta_k) + e_k,$$

where μ is the true overall mean, e_k is the *within-country* variation (with standard error s_k), independent of the variation of country means which is assumed to be $\zeta_k \sim N(0, \tau^2)$. This provides the marginal distributions $m_k \sim N(\mu, s_k^2 + \tau^2)$. In other words, the random-effect model assumes the variation in country means comes from not only sampling error, but also underlying population heterogeneity (*between-country* variation ζ_k).¹² Thus, our goal is to quantify the fraction of variation resulting from between-country heterogeneity.

To estimate the fraction, we first need to estimate τ^2 and then the true overall mean μ . We follow the approach of DerSimonian and Laird (1986) and adopt the method of moments to estimate $\hat{\tau}^2$. This is the simplest and most commonly used approach in meta-analysis. This method is essentially the method of moments which requires to calculate the Q statistic:

$$Q = \sum_{k=1}^K w_k \left[m_k - \frac{\sum_{k=1}^K w_k m_k}{\sum_{k=1}^K w_k} \right]^2,$$

where K is the number of countries in our data set and $w_k = \frac{1}{s_k^2}$. Under the assumption of random-effect model, we can obtain that

$$\mathbb{E}(Q) = \left(\sum_{k=1}^K w_k - \frac{\sum_{k=1}^K w_k^2}{\sum_{k=1}^K w_k} \right) (\tau^2 + s_k^2) = (K - 1) + \left(\sum_{k=1}^K w_k - \frac{\sum_{k=1}^K w_k^2}{\sum_{k=1}^K w_k} \right) \tau^2.$$

As we plug-in the moment condition, we can obtain

$$Q = (K - 1) + \left(\sum_{k=1}^K w_k - \frac{\sum_{k=1}^K w_k^2}{\sum_{k=1}^K w_k} \right) \hat{\tau}_{DL}^2.$$

¹²In the fixed-effect model, the variation is assumed to be purely from the sampling error, so the model reduces to $m_k = \mu + e_k$.

Rearranging the moment condition of τ^2 , we can obtain the DerSimonian and Laird estimate

$$\hat{\tau}_{DL}^2 = \frac{Q - (K - 1)}{\sum_{k=1}^K w_k - \frac{\sum_{k=1}^K w_k^2}{\sum_{k=1}^K w_k}}.$$

By convention, $\hat{\tau}_{DL}^2$ is replaced with 0 if $Q < K - 1$. Once we obtain $\hat{\tau}_{DL}^2$, we can therefore plug it into the other moment condition, $\mathbb{E} \left[\frac{\sum_{k=1}^K w_k^* m_k}{\sum_{k=1}^K w_k^*} \right] = \mu$ for $w_k^* = \frac{1}{s_k^2 + \tau^2}$, and obtain

$$\hat{\mu}_{DL} = \frac{\sum_{k=1}^K \hat{w}_k^* m_k}{\sum_{k=1}^K \hat{w}_k^*},$$

where $\hat{w}_k^* = \frac{1}{s_k^2 + \hat{\tau}_{DL}^2}$. Moreover, we can obtain the confidence intervals by using the approximation

$$\hat{\mu}_{DL} \sim N \left(\mu, \frac{1}{\sum_{k=1}^K \hat{w}_k^*} \right).$$

After obtaining the estimators of the between-country variance $\hat{\tau}_{DL}^2$ and the overall mean $\hat{\mu}_{DL}$, we can now further compute the fraction of between-country variation among observed variation. Intuitively, suppose the sampling errors in different countries are the same, i.e. $s_k = s$, then the unconditional variance of country k is given by

$$\text{var}(m_k) = \tau^2 + s^2.$$

Therefore, the fraction of variance that is due to between-country heterogeneity is

$$i^2 = \frac{\tau^2}{\tau^2 + s^2},$$

which is the true value that we want to estimate. If $i^2 > 0$, then the true means of different countries are indeed heterogeneous. Here we follow [Higgins and Thompson \(2002\)](#), defining the I^2 statistic to be of the form

$$I^2 = \frac{\hat{\tau}_{DL}^2}{\hat{\tau}_{DL}^2 + \hat{s}^2},$$

where \hat{s}^2 is an estimate of s^2 . In the general case, the sampling errors in different countries would be different. That is, the s^2 would be replaced with a set of sampling errors $\{s_1, s_2, \dots, s_K\}$. In this case, [Higgins and Thompson \(2002\)](#) suggest that we can plug in

$$\hat{s}^2 = \frac{K - 1}{\sum_{k=1}^K w_k - \frac{\sum_{k=1}^K w_k^2}{\sum_{k=1}^K w_k}}.$$

After defining the I^2 statistic, we can rewrite the definition and obtain the following form:

$$I^2 = \left(\frac{Q - (K - 1)}{Q} \right) \times 100\%,$$

which is set to be 0 if $Q < K - 1$. Intuitively, Q statistic reflects the total amount of observed variation, and $K - 1$ is the expectation of Q if there is no between-country heterogeneity ($\tau = 0$). Thus, $Q - (K - 1)$ is the amount of excess variation, and I^2 is the fraction of variation that is from between-country heterogeneity. Lastly, $K - 1$ is actually the degree of freedom (df). As we replace $K - 1$ with df , we can obtain the form of I^2 in [Higgins et al. \(2003\)](#):

$$I^2 = \left(\frac{Q - df}{Q} \right) \times 100\%.$$

C.1 Proposal Offer

Table S8: Proposal Offers in Different Countries

| Countries | Mean | [95% C.I.] | | Weight (%) |
|----------------|-------|-------------|-------|------------|
| Spain | 27.73 | 25.42 | 30.04 | 10.36 |
| China | 30.83 | 26.16 | 35.50 | 5.90 |
| US Midwest | 35.18 | 33.73 | 36.64 | 12.18 |
| US North East | 35.20 | 34.01 | 36.38 | 12.66 |
| Singapore | 35.50 | 30.06 | 40.94 | 4.90 |
| England | 35.92 | 33.07 | 38.77 | 9.19 |
| US South | 36.19 | 35.28 | 37.09 | 13.10 |
| Japan | 36.64 | 32.35 | 40.93 | 6.48 |
| Canada | 37.30 | 35.89 | 38.71 | 12.26 |
| US West | 37.35 | 36.35 | 38.35 | 12.96 |
| Overall Effect | 35.00 | 33.50 | 36.51 | 100.00 |

Table S9: Heterogeneity Measures of Proposal Offers

| Q | df | p-value | $I^2(\%)$ | [95% C.I.] | | $\hat{\tau}^2$ |
|-------|----|---------|-----------|-------------|-------|----------------|
| 67.20 | 9 | 0.000 | 86.61 | 77.31 | 92.10 | 4.26 |

The results of Table S8 and S9 show that a significant amount of between-country heterogeneity exists in proposal offers ($p < 0.001$). In particular, 86.61% of the variation comes from between-country heterogeneity, and the estimated variance of the true effect size is 4.26.

C.2 Acceptance Rate

To compare conditional acceptance rates, we employ the estimated coefficients of the following logistic regression:

$$g(y_i) = \alpha \cdot \text{offer}_i + \sum_{j=1}^{10} \beta_j D_{ji} + \epsilon_i$$

where $g(\cdot)$ is the logit function and D_{ji} is the dummy variable of each countries. After running the regression, we compare the heterogeneity in the estimations of country dummy variables. Notice that the standard errors are calculated by clustering at the session level.

The acceptance rate results of Table S10 and S11 show insignificant between-country heterogeneity ($p = 0.164$). In fact, only 30.63% of the variation comes from between-country heterogeneity, and the estimated variance of the true effect size is only 0.012.

Table S10: Estimations of Country Dummy Variables in the Logistic Regression

| Country | Mean | [95% C.I.] | | Weight (%) |
|----------------|-------|-------------|-------|------------|
| England | -1.79 | -2.09 | -1.50 | 11.61 |
| US West | -1.74 | -2.00 | -1.48 | 13.65 |
| Canada | -1.57 | -1.88 | -1.26 | 10.96 |
| Spain | -1.57 | -1.97 | -1.17 | 7.40 |
| Japan | -1.57 | -1.96 | -1.18 | 7.75 |
| US South | -1.52 | -1.75 | -1.30 | 15.88 |
| US North East | -1.43 | -1.69 | -1.17 | 13.37 |
| US Midwest | -1.29 | -1.58 | -0.99 | 11.65 |
| China | -1.14 | -1.76 | -0.52 | 3.53 |
| Singapore | -1.08 | -1.64 | -0.52 | 4.21 |
| Overall Effect | -1.52 | -1.65 | -1.40 | 100.00 |

Table S11: Heterogeneity Measures of Acceptance Rates

| Q | df | p-value | $I^2(\%)$ | [95% C.I.] | | $\hat{\tau}^2$ |
|-------|----|---------|-----------|-------------|-------|----------------|
| 12.97 | 9 | 0.164 | 30.63 | 0.00 | 66.85 | 0.012 |

C.3 First Transaction

Table S12: First Standardized Price Difference in Different Countries

| Country | Mean | [95% C.I.] | | Weight (%) |
|----------------|-------|-------------|------|------------|
| Spain | -0.22 | -0.65 | 0.20 | 0.40 |
| Singapore | -0.15 | -0.73 | 0.44 | 0.22 |
| US South | -0.02 | -0.05 | 0.00 | 31.06 |
| US West | -0.02 | -0.07 | 0.02 | 21.15 |
| England | -0.01 | -0.14 | 0.13 | 3.91 |
| US North East | 0.01 | -0.05 | 0.07 | 13.17 |
| China | 0.04 | -0.11 | 0.18 | 3.22 |
| Canada | 0.04 | -0.01 | 0.09 | 18.61 |
| Japan | 0.08 | -0.17 | 0.32 | 1.21 |
| US Midwest | 0.09 | -0.01 | 0.18 | 7.04 |
| Overall Effect | 0.00 | -0.02 | 0.03 | 100.00 |

Table S13: Heterogeneity Measures of First Standardized Price Difference

| Q | df | p-value | $I^2(\%)$ | [95% C.I.] | | $\hat{\tau}^2$ |
|-------|----|---------|-----------|-------------|-------|----------------|
| 12.38 | 9 | 0.193 | 27.28 | 0.00 | 65.05 | 0.00 |

Table S12 and S13 show that between-country heterogeneity in the first transaction price is not significant ($p = 0.193$). Moreover, most of the variation comes from within-country variance—only 27.28% of the variation is from between-country heterogeneity. In fact, the overall average is not significantly different from 0, ranging from -22% to 9% . Lastly, the estimated variance of the true effect size is less than 0.01.

C.4 Last Transaction

The last transaction results of Table S14 and S15 show that the between-country heterogeneity is also not significant ($p = 0.276$). Most of the variation is still from within-country variance—only 18.16% of the variation comes from between-country heterogeneity, with mean effects ranging from -2% to 12% . Lastly, the estimated variance of the true effect size is also less than 0.01.

Table S14: Last Standardized Price Difference in Different Countries

| Country | Mean | [95% C.I.] | | Weight (%) |
|----------------|-------|-------------|-------|------------|
| US South | -0.02 | -0.04 | -0.00 | 34.17 |
| US West | -0.01 | -0.04 | 0.01 | 22.74 |
| Singapore | 0.00 | -0.27 | 0.27 | 0.28 |
| US North East | 0.00 | -0.03 | 0.04 | 15.88 |
| England | 0.02 | -0.08 | 0.11 | 2.36 |
| China | 0.02 | -0.04 | 0.09 | 4.35 |
| Canada | 0.02 | -0.01 | 0.06 | 14.83 |
| US Midwest | 0.03 | -0.04 | 0.10 | 4.52 |
| Spain | 0.13 | -0.20 | 0.45 | 0.20 |
| Japan | 0.13 | -0.05 | 0.30 | 0.68 |
| Overall Effect | -0.00 | -0.02 | 0.01 | 100.00 |

Table S15: Heterogeneity Measures of Last Standardized Price Difference

| Q | df | p-value | $I^2(\%)$ | [95% C.I.] | | $\hat{\tau}^2$ |
|-------|----|---------|-----------|-------------|-------|----------------|
| 11.00 | 9 | 0.276 | 18.16 | 0.00 | 59.08 | 0.00 |

D Games

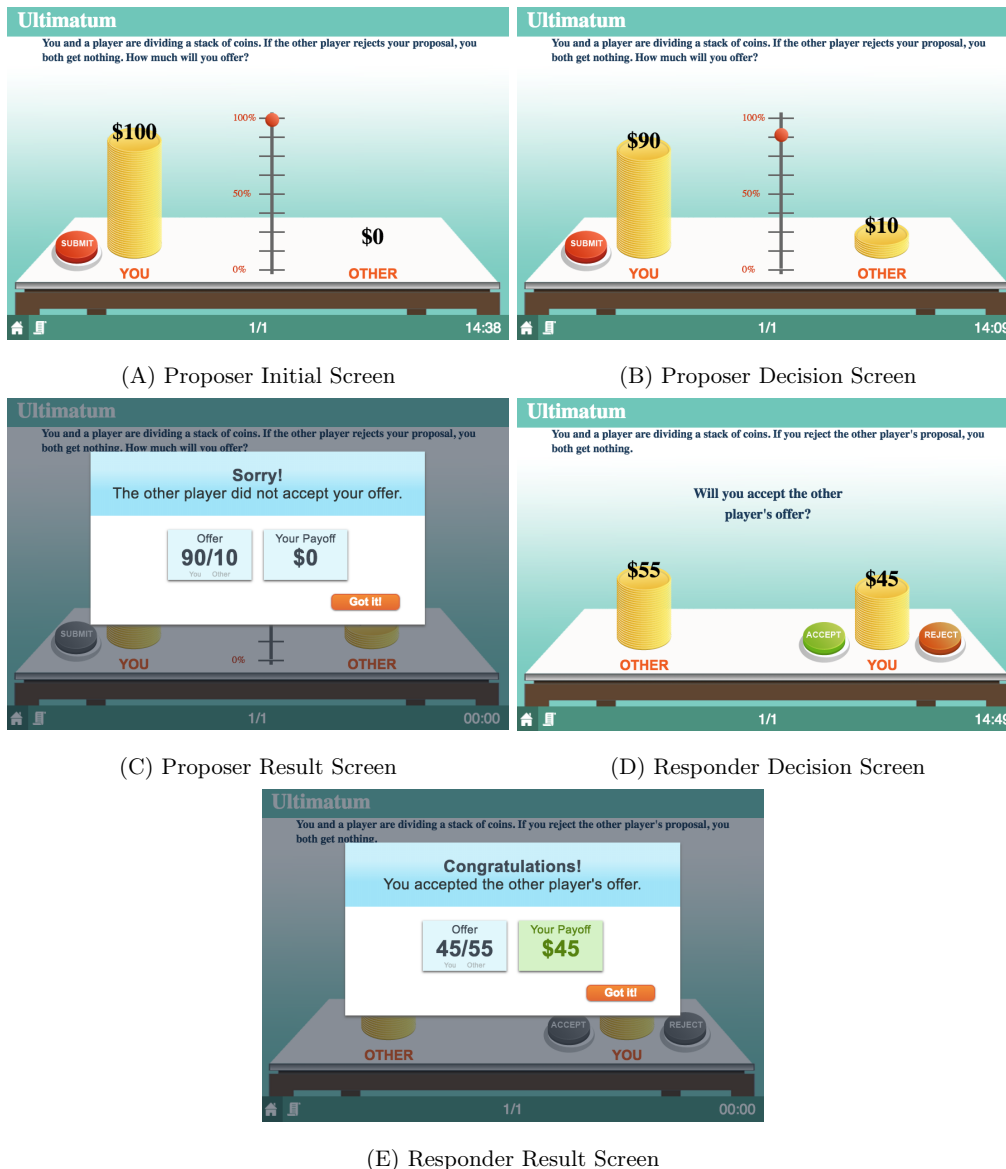


Figure S7: Screenshots in each stage of the ultimatum game

In this section, we provide the screenshots of each stage in our ultimatum game and buyer-seller double auction. Figure S7 shows screenshots in the ultimatum game. The Proposers' screenshots are S7A to S7C. At the beginning of the game, the Proposer would see the amount of money that he can split with the Responder. The Proposer then can drag the ball in the middle of the screen to make the proposal. After the Responder makes a decision, the Proposer would see the result screen. Figure S7C is a sample result screen that the player sees if the proposal offer is rejected.

The Responder's screenshots are shown in Figure S7D and S7E. After the Proposer makes the proposal offer, the Responder would see the proposed allocation and decides whether to accept or

reject it through the buttons on the screen. After the decision is made, the players would see the result screen and Figure S7E is a sample result screen when the proposal offer is accepted.



Figure S8: Screenshots in each stage of the buyer-seller double auction

In the buyer-seller double auction, the buyers' initial screenshot is shown in Figure S8A and the sellers' initial screenshot is in S8D. Both buyers and sellers would see their values/costs of the commodities as they enter the market. Then, buyers and sellers can trade in the market by either dragging the slider bar to make an offer or clicking the button to accept the standing offer (Figure S8B for buyers). Notice that if a player is making a potential trade that would give them negative payoff, the slider bar would turn red as a warning signal (Figure S8E for sellers). After a transaction occurs, the player would see the result screen which shows the price and profit of that transaction. Figure S8C and Figure S8F are the result screens for buyers and sellers, respectively.